

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 220.098 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-5

Perfect score: 20

Sequence: 1 tggagcggcagtggtgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sv.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg\_hum.\*  
31: em.htg\_inv.\*  
32: em.htg\_other.\*  
33: em.htg\_mus.\*  
34: em.htg\_pln.\*  
35: em.htg\_rtd.\*  
36: em.htg\_mam.\*  
37: em.htg\_vrt.\*  
38: em\_sy.\*  
39: em.htgo\_hum.\*  
40: em.htgo\_mus.\*  
41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	20	100.0	20	6	AX014705	Sequence
2	20	100.0	2765	6	AX014701	AX014701 Sequence
c 3	20	100.0	174953	2	AC094732	AC094732 Rattus no
4	18.4	92.0	2583	10	AF302075	AF302075 Mus muscu
5	18.4	92.0	2601	10	AF157106	AF157106 Mus muscu
6	18.4	92.0	2652	10	AF302076	AF302076 Mus muscu
7	18.4	92.0	2694	10	AF302077	AF302077 Mus muscu
8	18.4	92.0	2892	10	AF157105	AF157105 Mus muscu
9	18.4	92.0	2925	6	AX033272	AX033272 Sequence
10	18.4	92.0	203286	10	AF176569	AF176569 Mus muscu
11	18.4	92.0	203286	10	AF176569	AF176569 Mouse DNA
12	18.4	92.0	208249	2	AL607032	AL607032 Mus muscu
13	17.4	87.0	25478	2	AC007116	AC007116 Drosophil
c 14	17.4	87.0	82665	2	AC006938	AC006938 Drosophil
15	17.4	87.0	110000	2	AC096785_1	Continuation (2 of
c 16	17.4	87.0	117812	2	AC020493	AC020493 Drosophil
17	17.4	87.0	130935	2	AP004192	AP004192 Oryza sat
18	17.4	87.0	135085	2	AC025620	AC025620 Homo sapi
19	17.4	87.0	136485	9	AC007422	AC007422 Homo sapi
c 20	17.4	87.0	159998	2	AC074292	AC074292 Homo sapi
c 21	17.4	87.0	166973	9	AC084117	AC084117 Homo sapi
22	17.4	87.0	169850	2	AC109619	AC109619 Mus muscu
23	17.4	87.0	172826	3	AC008353	AC008353 Drosophil
24	17.4	87.0	179779	2	AP005315	AP005315 Oryza sat
25	17.4	87.0	196410	9	AC027544	AC027544 Homo sapi
26	17.4	87.0	241174	2	AC116107	AC116107 Mus muscu
c 27	17.4	87.0	242454	2	AC091342	AC091342 Rattus no
28	17.4	87.0	298827	3	AE003463	AE003463 Drosophil
c 29	16.8	84.0	558	14	AF246000	AF246000 Hepatitis
30	16.8	84.0	586	3	NT066265	U66265 Nicotiana t
c 31	16.8	84.0	915	3	FJDNAIVC	X95816 P.jarrei co
c 32	16.8	84.0	1174	9	HSAT9121	AJ279121 Homo sapi
c 33	16.8	84.0	2130	9	AK057871	AK057871 Homo sapi
34	16.8	84.0	2489	1	ECFTSQA	X02821 E.coli cel
35	16.8	84.0	3248	8	AB040136	AB040136 Porphyrid
36	16.8	84.0	3254	8	AB040135	AB040135 Porphyrid
37	16.8	84.0	3333	1	ECOF7SQA	X02668 E.coli ddi
c 38	16.8	84.0	4272	9	AF298151	AF298151 Homo sapi
c 39	16.8	84.0	4273	9	AF298152	AF298152 Homo sapi
c 40	16.8	84.0	5893	3	AY050232	AY050232 Drosophil
c 41	16.8	84.0	7207	9	HSAT38520	AJ238520 Homo sapi
c 42	16.8	84.0	9410	14	HPCKIR2	D50481 Hepatitis C
c 43	16.8	84.0	9410	14	HPCKIR2	D50485 Hepatitis C
c 44	16.8	84.0	12518	1	AE005186	AE005186 Escherich
c 45	16.8	84.0	13159	2	AC023744	AC023744 Drosophil

# ALIGNMENTS

RESULT 1  
LOCUS AX014705  
DEFINITION Sequence 5 from Patent WO9953077.  
ACCESSION AX014705  
VERSION AX014705.1 GI:10040978  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Schwartz J.C., Gros, C., Ouimet, T., Rose, C., Bonhomme, M.C. and  
TITLE Novel nep II membrane metalloprotease and its use for screening  
inhibitors useful in therapy

JOURNAL Patent: WO 9953077-A 5 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
Location/Qualifiers

FEATURES  
source  
1..20  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
2 a 3 c 10 g 5 t

BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
|||||  
1 TGGAGCGGCAGTTGGCTGTG 20

Db 1 TGGAGCGGCAGTTGGCTGTG 20  
|||||  
1 TGGAGCGGCAGTTGGCTGTG 20

RESULT 2  
AX014701  
LOCUS  
DEFINITION  
Sequence 1 from Patent WO9953077.  
ACCESSION  
AX014701  
VERSION  
AX014701.1 GI:10040975  
KEYWORDS  
SOURCE  
ORGANISM  
black rat.  
Rattus rattus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
Schwartz J.C., Gros C., Oumet T., Rose C., Bonhomme M.C. and  
Facchinetti P.  
TITLE  
Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
JOURNAL  
Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
Location/Qualifiers

FEATURES  
source  
1..2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107..2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSSVGMERADNCGRRRLGFEVCGLLVLLTLLMGAIVTL  
GVFYSIGKQLPLNLSLHVSRHRTVYKRVLRDSSQKSDICTPSCVIAAARIQNDP  
QSKPCDNFQVACGGLRHVHPETNSRYSDPDLRDEVLKGLVEDSSVQHRPA  
VEKATLRYSCMNQSVLEKRDSEPLNVLDMIGGWPVMDKKNWETGPKVELERQLAV  
LNSQFNRRVLDLIFWDDQNSRRHYIIDPTIGMPSRETYFKEDSHRVREAYLQFM  
LVRLVLDRLNLPGETDLVQEEAAQVHLHETLANATVPQFGRHDVLTALYHRMGLFE  
LQRFGLKGFNWTFLFQNLVSSVQVELLPNEVVVYGPYLEEIIIDVFPAQTLQ  
IKPAFSKDSKSIYSELIEKIRSVFVNDLDELNMWDESKKAEKALINIEQIGYPDY  
ILEDNRHLDDEVSLLFSDLLFENGLQNLKNAQSLKLRKVQDNWLGIAAVY  
NAFYSNNRNLVFPAGLQPPFSKQDQALNFGGGMVIGHEITHGFDNDGRNFRDK  
GNMLDWNFSARHFRQSQCMQIYQSNFELADNQNVNGFTLGENIADNGVROA  
YKAYLQWAGGRDQLRGLNLTAYQFLFNAYQVCGSGYRPFQIAQSIKTDVHSPILK  
YRVLSGLQNLPGSFQAFHCPRGSPMHPMNCRIW"

BASE COUNT  
ORIGIN  
684 a 735 c 787 g 559 t

Query Match 100.0%; Score 20; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20

Db 1 TGGAGCGGCAGTTGGCTGTG 20

RESULT 3  
AC094732/c  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-516, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
56 unordered pieces.  
AC094732  
HTG: HTGS PHASE1.  
VERSION  
AC094732.2 GI:17941511  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS  
Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,  
Alsbrooks S.L., Amarantunga H.C., Are J.R., Banks T., Barbara J.,  
Benton J., Blmage K., Blankenburg K., Bonnin D., Bouck J.,  
Bowie S., Brieva M., Brown E., Brown M., Bryant N.P., Buhay C.,  
Burch P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,  
Carter M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,  
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,  
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,  
Davy-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,  
Denn A.L., Ding Y., Dinh H.H., Douthwaite K.J., Draper H.,  
Dugan-Rocha S., Durbin K.J., Earnhart C., Edgar D., Edwards C.C.,  
Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,  
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,  
Garza N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,  
Hamilton K., Harris C., Harris K., Hart M., Havlak P., Hawes A.,  
Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C.,  
Hollins B., Homs F., Howard S., Huber J., Hulyk S., Hume J.,  
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jollivet S.,  
Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,  
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,  
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W.,  
Louiseged H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,  
Ma J., Maheshwari M., Mapua P., Martin R., Martindale A.,  
Martinez E., Massey E., Mawhiney E., McLeod M.P., Meador M.,  
Mei G., Metzker M., Miner G., Miner Z., Mitchell T., Mohabbat K.,  
Morgan M., Morris S., Moser M., Neal D., Newton J., Newton N.,  
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokenwo S.,  
Ogih M., Okwono G., Oragunye N., Oviedo R., Pace A., Payton B.,  
Peery J., Perez L., Peters L., Pickens R., Primus E., Pu L.L.,  
Quiles M., Ren Y., Rives M., Rojas A., Rojibokan I., Rolfe M.,  
Ruiz S., Savary G., Scherer S., Scott G., Shen H., Shoostari N.,  
Stinson I., Sodergren E., Sonalke T., Sparks A., Stanley H.,  
Stone H., Sutton A., Svatek A., Tabor P., Tamerisa A., Tamerisa K.,  
Tang H., Tansey J., Taylor C., Taylor T., Telford B., Thomas N.,  
Thomas S., Usmani K., Vasquez L., Vera V., Villalon D., Vinson R.,  
Wall R., Wang S., Ward-Moore S., Warren R., Washington C.,  
Watlington S., Williams G., Williamson A., Wlaczky R., Wooden S.,  
Worley K., Wu C., Wu Y., Wu Y.F., Zhou J., Zorrilla S., Nelson D.,  
Weinstock G. and Gibbs R.

TITLE  
Direct Submission  
JOURNAL  
2 (bases 1 to 174953)  
REFERENCE  
Worley K.C.  
AUTHORS  
TITLE  
JOURNAL  
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15624568.

COMMENT  
Center: Baylor College of Medicine  
Genome Center  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: GHGF

Center clone name: CH230-516  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to  
findPhrapList.  
Consensus quality: 152255 bases at least Q40  
Consensus quality: 158448 bases at least Q30  
Consensus quality: 164461 bases at least Q20  
Estimated insert size: 155965; sum-of-coverage estimation.  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.2x in Q20 bases; sum-of-coverage estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 56 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 13782: contig of 13782 bp in length  
13783 13882: gap of unknown length  
13883 23287: contig of 9405 bp in length  
23288 23387: gap of unknown length  
23388 28081: contig of 4694 bp in length  
28082 28181: gap of unknown length  
28182 33807: contig of 5626 bp in length  
33808 33907: gap of unknown length  
33908 39271: contig of 5364 bp in length  
39272 39371: gap of unknown length  
39372 44270: contig of 4899 bp in length  
44271 44370: gap of unknown length  
44371 47723: contig of 3353 bp in length  
47724 47823: gap of unknown length  
47824 53427: contig of 5604 bp in length  
53428 53527: gap of unknown length  
53528 57303: contig of 3776 bp in length  
57304 57403: gap of unknown length  
57404 62018: contig of 4615 bp in length  
62019 62118: gap of unknown length  
62119 66676: contig of 4558 bp in length  
66677 66776: gap of unknown length  
66777 70201: contig of 3425 bp in length  
70202 70301: gap of unknown length  
70302 74082: contig of 3781 bp in length  
74083 74182: gap of unknown length  
74183 78306: contig of 4124 bp in length  
78307 78406: gap of unknown length  
78407 81222: contig of 2816 bp in length  
81223 81322: gap of unknown length  
81323 84552: contig of 3230 bp in length  
84553 84652: gap of unknown length  
84653 88839: contig of 4187 bp in length  
88840 88939: gap of unknown length  
88940 92256: contig of 3317 bp in length  
92257 92356: gap of unknown length  
92357 94995: contig of 2639 bp in length  
94996 95095: gap of unknown length  
95096 98465: contig of 3370 bp in length  
98466 98565: gap of unknown length  
98566 101632: contig of 3067 bp in length  
101633 101732: gap of unknown length  
101733 104939: contig of 3207 bp in length  
104940 105039: gap of unknown length  
105040 109248: contig of 4209 bp in length  
109249 109348: gap of unknown length  
109349 112435: contig of 3087 bp in length  
112436 112536: gap of unknown length  
112536 114676: contig of 2141 bp in length  
114677 114776: gap of unknown length  
114777 117567: contig of 2791 bp in length  
117568 117667: gap of unknown length

117668 119209: contig of 1542 bp in length  
119210 119309: gap of unknown length  
119310 121797: contig of 2488 bp in length  
121798 121897: gap of unknown length  
121898 124293: contig of 2396 bp in length  
124294 124393: gap of unknown length  
124394 126428: contig of 2035 bp in length  
126429 126528: gap of unknown length  
126529 129325: contig of 2997 bp in length  
129326 129625: gap of unknown length  
129626 132334: contig of 2709 bp in length  
132335 132434: gap of unknown length  
132435 135274: contig of 2840 bp in length  
135275 135374: gap of unknown length  
135376 138074: contig of 2700 bp in length  
138075 138174: gap of unknown length  
138176 139985: contig of 1811 bp in length  
139986 140085: gap of unknown length  
140086 142273: contig of 2188 bp in length  
142274 142373: gap of unknown length  
142374 143598: contig of 1225 bp in length  
143599 143698: gap of unknown length  
143700 145434: contig of 1736 bp in length  
145435 145534: gap of unknown length  
145535 146985: contig of 1451 bp in length  
146986 147085: gap of unknown length  
147086 148099: contig of 1014 bp in length  
148100 148199: gap of unknown length  
148200 150915: contig of 2716 bp in length  
150916 151015: gap of unknown length  
151016 152501: contig of 1486 bp in length  
152502 152601: gap of unknown length  
152602 154010: contig of 1409 bp in length  
154011 154110: gap of unknown length  
154111 155758: contig of 1648 bp in length  
155759 155858: gap of unknown length  
155859 157622: contig of 1764 bp in length  
157623 157722: gap of unknown length  
157723 159428: contig of 1706 bp in length  
159429 159528: gap of unknown length  
159529 161209: contig of 1681 bp in length  
161210 161309: gap of unknown length  
161310 163413: contig of 2104 bp in length  
163414 163513: gap of unknown length  
163514 164702: contig of 1189 bp in length  
164703 164802: gap of unknown length  
164803 165998: contig of 1196 bp in length  
165999 166098: gap of unknown length  
166099 167412: contig of 1314 bp in length  
167413 167512: gap of unknown length  
167513 169231: contig of 1719 bp in length  
169232 169331: gap of unknown length  
169332 170534: contig of 1203 bp in length  
170535 170634: gap of unknown length  
170635 172047: contig of 1413 bp in length  
172048 172147: gap of unknown length  
172148 173509: contig of 1362 bp in length  
173510 173610: gap of unknown length  
173610 174953: contig of 1344 bp in length.

## FEATURES

Location/Qualifiers

Query Match 100.0%; Score 20; DB 2; Length 174953;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
|||||  
Db 60956 TGGAGCGGCAGTTGGCTGTG 60937

RESULT 4  
AF302075  
LOCUS AF302075 2583 bp mRNA linear ROD 11-JUN-2001

```

DEFINITION Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.
ACCESSION AF302075
VERSION AF302075.1 GI:10505359
KEYWORDS Mus musculus.
SOURCE Mus musculus
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2583)
AUTHORS Shirotani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,
  Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,
  Iwatsubo,T. and Saido,T.C.
TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most
  rapidly and efficiently among thiorphan- and
  phosphoramidon-sensitive endopeptidases
JOURNAL J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE 21293028
PUBMED 11278416
REFERENCE 2 (bases 1 to 2583)
AUTHORS Shirotani,K. and Saido,T.C.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain
  Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama
  351-0190, Japan
FEATURES
  source
    1. .2583
    /organism="Mus musculus"
    /db_xref="taxon:10090"
  CDS
    25. ..2253
    /note="endopeptidase"
    /codon_start=1
    /product="neprilysin-like peptidase alpha"
    /protein_id="AAG18446.1"
    /db_xref="GI:10505360"
    /translation="MYERAGWCKKSPGFVEYGLMVLVLLLLGAIIVTLGVVYSIALRD
    SSKSDICTPSCVIAAARILENMQSRNCFYQACGGLRHVHPETNSRYSVF
    DILRDEVLKGVLEDSOHRPAVEKAKTLVRSQNSQVIEKRDSEPLLSVLKMGV
    GWPVLDKKNETGLKWELEROLAVLNSQNRRLVLDLFWNDQNSRHVVIIDOPT
    LGMPREYFQEDNNHKVRKAYLEFMTSVATMLRKQNLKSKESAMVREMAEVLLET
    HLANATVPQEKRDHVTALYHRMDLMEQLQERGLKGFNTLFIQNLVSSVEVLPDDE
    VVYGIPLYLENDIIDSYSARTMQLVWRLVLDLIGLSLQRFKARVDYRKALYGT
    TWDEESKKAQEKAMNRIQIGYDYLENNKHLEDEYSLSLTFEDLYPENGQNLK
    NNAQRLSKLREKVDONLWIGAAVNAFSPNRNOIVPAGILOPFPSKQDQPSLN
    FGGICNVGHEITHGDDNGRFDKNGMLDWSNFSARHFQOOSQCMIIQYGNFSWE
    LADNQVNGFSTLGENIADNGVRQAYKATLRLWLDGKQDQRLPGLNLTAAQLFFIN
    AQWCGSYRPEFAVQSIKTDVHSPKLYRVLGSLQNLPGFSEAFHCPRGSPMPMKRCR
    IW"
  BASE COUNT 665 a 667 c 736 g 515 t
  ORIGIN
    Query Match 92.0%; Score 18.4; DB 10; Length 2583;
    Best Local Similarity 95.0%; Pred. No. 1.1e+02;
    Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 TGGAGCGCGAGTTGGCTGTG 20
    ||||| ||||| ||||| |||||
  Db 560 TGGAGCGACAGTTGGCTGTG 579

RESULT 5
LOCUS AF157106 2601 bp mRNA linear ROD 25-NOV-1999
DEFINITION Mus musculus soluble secreted endopeptidase delta mRNA,
  alternatively spliced product., complete cds.
ACCESSION AF157106
VERSION AF157106.1 GI:6467400
KEYWORDS Mus musculus.
SOURCE Mus musculus
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2601)

DEFINITION Mus musculus neprilysin-like peptidase beta mRNA, complete cds.
ACCESSION AF302076
VERSION AF302076.1 GI:10505361
KEYWORDS Mus musculus.
SOURCE Mus musculus
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2652)
AUTHORS Shirotani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,
  Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,
  Iwatsubo,T. and Saido,T.C.
TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most
  rapidly and efficiently among thiorphan- and
  phosphoramidon-sensitive endopeptidases
JOURNAL J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE 21293028
PUBMED 11278416
REFERENCE 2 (bases 1 to 2652)
AUTHORS Shirotani,K. and Saido,T.C.

```

```

AUTHORS Ikeda,K., Emoto,N., Raharjo,S.B., Nurhantari,Y., Saiki,K.,
  Yokoyama,M. and Matsuo,M.
TITLE Molecular identification and characterization of novel
  membrane-bound metalloprotease, the soluble secreted form of which
  hydrolyzes a variety of vasoactive peptides
JOURNAL J. Biol. Chem. 274 (45), 32469-32477 (1999)
MEDLINE 20011457
PUBMED 10542292
REFERENCE 2 (bases 1 to 2601)
AUTHORS Ikeda,K., Emoto,N. and Matsuo,M.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (08-JUN-1999) International Center for Medical Research,
  Kobe University School of Medicine, 7-5-1 Kusunoki, Chuo, Kobe
  6500017, Japan
FEATURES
  source
    1. .2601
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    59. .2287
    /note="SEP(delta); metalloprotease; alternatively spliced"
    /codon_start=1
    /product="soluble secreted endopeptidase delta"
    /protein_id="AAF13153.1"
    /db_xref="GI:6467401"
    /translation="MYERAGWCKKSPGFVEYGLMVLVLLLLGAIIVTLGVVYSIALRD
    SSKSDICTPSCVIAAARILENMQSRNCFYQACGGLRHVHPETNSRYSVF
    DILRDEVLKGVLEDSOHRPAVEKAKTLVRSQNSQVIEKRDSEPLLSVLKMGV
    GWPVLDKKNETGLKWELEROLAVLNSQNRRLVLDLFWNDQNSRHVVIIDOPT
    LGMPREYFQEDNNHKVRKAYLEFMTSVATMLRKQNLKSKESAMVREMAEVLLET
    HLANATVPQEKRDHVTALYHRMDLMEQLQERGLKGFNTLFIQNLVSSVEVLPDDE
    VVYGIPLYLENDIIDSYSARTMQLVWRLVLDLIGLSLQRFKARVDYRKALYGT
    TWDEESKKAQEKAMNRIQIGYDYLENNKHLEDEYSLSLTFEDLYPENGQNLK
    NNAQRLSKLREKVDONLWIGAAVNAFSPNRNOIVPAGILOPFPSKQDQPSLN
    FGGICNVGHEITHGDDNGRFDKNGMLDWSNFSARHFQOOSQCMIIQYGNFSWE
    LADNQVNGFSTLGENIADNGVRQAYKATLRLWLDGKQDQRLPGLNLTAAQLFFIN
    AQWCGSYRPEFAVQSIKTDVHSPKLYRVLGSLQNLPGFSEAFHCPRGSPMPMKRCR
    IW"
  BASE COUNT 655 a 681 c 748 g 517 t
  ORIGIN
    Query Match 92.0%; Score 18.4; DB 10; Length 2601;
    Best Local Similarity 95.0%; Pred. No. 1.1e+02;
    Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 TGGAGCGCGAGTTGGCTGTG 20
    ||||| ||||| ||||| |||||
  Db 594 TGGAGCGACAGTTGGCTGTG 613

RESULT 6
LOCUS AF302076 2652 bp mRNA linear ROD 11-JUN-2001
DEFINITION Mus musculus neprilysin-like peptidase beta mRNA, complete cds.
ACCESSION AF302076
VERSION AF302076.1 GI:10505361
KEYWORDS Mus musculus.
SOURCE Mus musculus
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 2652)
AUTHORS Shirotani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,
  Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,
  Iwatsubo,T. and Saido,T.C.
TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most
  rapidly and efficiently among thiorphan- and
  phosphoramidon-sensitive endopeptidases
JOURNAL J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE 21293028
PUBMED 11278416
REFERENCE 2 (bases 1 to 2652)
AUTHORS Shirotani,K. and Saido,T.C.

```

**TITLE** Direct Submission  
**JOURNAL** Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan  
**FEATURES** Location/Qualifiers  
 source 1. .2652  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 CDS 25. .2322  
 /note="endopeptidase"  
 /codon\_start=1  
 /product="neprilysin-like peptidase beta"  
 /protein\_id="AAG18447.1"  
 /db\_xref="GI:10505362"  
 /translation="MVERAGWCRKSPGFVEYGLMVLVLLLLGAIIVTLGVFYSIGKOL PLTSLHFSWDEITVVKRALRDSLSKSDICTTPSCVIAAARILENMQSRNCPENFY QYACGWLHRIHVIPETNSRYSVFDILRDELEVILKGVLEDSSTQHRPAVEKAKTLVRSQMNSQVIEKRDSEPLLSVLMKVGWPMVAMDKWNETMGLKWELEROLAVLNSQFNRRVL IDLFINWDDQSSRRHVIYIDQPTLGMPSREYFQEDNNHKVKAYLEFMTSVATMLRK DONLSKESAMVREEMAEVLELEHLANATVQEKRDHVTALYHRMDLMELOERFLGK FNWTLFTQNLSSVEVLEFPDEEVVVGIPYLENLEDIIDSYSAPFTMONYLVRFLVD RIGLSQREFKARDYRKALYGTTFVEVRWRECVSVNMSAVSGSLYIKRAFSDKS KSTVRELIEKRSFVNDLDELNMDESKKAKQEKAMNIREQIGPYDILFEDNNKHL DEEYSITFTFDLYFENGQNLKNNQARSLKREKVDQNLMIIGAAVYNVAFYSNRRN QIVFPAGLPPFFSKDQPOQSLNFGGIGWIGHEITHGFDNNGRFGKNNMLDWSN FSAHFQOOSCMYQYGNFSEWELADNONVNGESTLGENIADNGGVRQAYKAYLRWLA DGGKDLRLGLNLTAQOLFFINVAQWCGSYRPEFAVOSIKTDVHSPKRYRVLGSLQN LPGFSEAFHCPRGSPMPMKRCRIW"  
 BASE COUNT 682 a 685 c 755 g 530 t  
 ORIGIN  
 Query Match 92.0%; Score 18.4; DB 10; Length 2652;  
 Best Local Similarity 95.0%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TGGAGCGGCAGTGGCTGTG 20  
 ||||| ||||| ||||| |||||  
 Db 629 TGGAGCGCAGTGGCTGTG 648  
 RESULT 7  
 LOCUS AF302077 2694 bp mRNA linear ROD 11-JUN-2001  
 DEFINITION Mus musculus neprilysin-like peptidase gamma mRNA, complete cds.  
 ACCESSION AF302077  
 VERSION AF302077.1 GI:10505363  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2694)  
 AUTHORS Shirotani, K., Tsubuki, S., Iwata, N., Takaki, Y., Harigaya, W., Maruyama, K., Kiryu-Seo, S., Kiyama, H., Iwata, H., Tomita, T., Iwatsubo, T. and Saido, T.C.  
 TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most rapidly and efficiently among thiorphan- and phosphoramidon-sensitive endopeptidases  
 JOURNAL J. Biol. Chem. 276 (24), 21895-21901 (2001)  
 MEDLINE 21293028  
 PUBMED 11278416  
 REFERENCE 2 (bases 1 to 2694)  
 AUTHORS Shirotani, K. and Saido, T.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan  
 LOCATION/Qualifiers  
 source 1. .2694  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 CDS 25. .2364  
 /note="endopeptidase"  
 /codon\_start=1  
 /product="neprilysin-like peptidase gamma"  
 /protein\_id="AAG18448.1"  
 /db\_xref="GI:10505364"  
 /translation="MVERAGWCRKSPGFVEYGLMVLVLLLLGAIIVTLGVFYSIGKOL PLTSLHFSWDEITVVKRALRDSLSKSDICTTPSCVIAAARILENMQSRNCPENFY QYACGWLHRIHVIPETNSRYSVFDILRDELEVILKGVLEDSSTQHRPAVEKAKTLVRS QYACGWLHRIHVIPETNSRYSVFDILRDELEVILKGVLEDSSTQHRPAVEKAKTLVRS CMNQSQVIEKRDSEPLLSVLMKVGWPMVAMDKWNETMGLKWELEROLAVLNSQFNRRVL IDLFINWDDQSSRRHVIYIDQPTLGMPSREYFQEDNNHKVKAYLEFMTSVATMLRK DONLSKESAMVREEMAEVLELEHLANATVQEKRDHVTALYHRMDLMELOERFLGK FNWTLFTQNLSSVEVLEFPDEEVVVGIPYLENLEDIIDSYSAPFTMONYLVRFLVD

/codon\_start=1  
 /product="neprilysin-like peptidase gamma"  
 /protein\_id="AAG18448.1"  
 /db\_xref="GI:10505364"  
 /translation="MVERAGWCRKSPGFVEYGLMVLVLLLLGAIIVTLGVFYSIGKOL PLTSLHFSWDEITVVKRALRDSLSKSDICTTPSCVIAAARILENMQSRNCPENFY QYACGWLHRIHVIPETNSRYSVFDILRDELEVILKGVLEDSSTQHRPAVEKAKTLVRSQMNSQVIEKRDSEPLLSVLMKVGWPMVAMDKWNETMGLKWELEROLAVLNSQFNRRVL IDLFINWDDQSSRRHVIYIDQPTLGMPSREYFQEDNNHKVKAYLEFMTSVATMLRK DONLSKESAMVREEMAEVLELEHLANATVQEKRDHVTALYHRMDLMELOERFLGK FNWTLFTQNLSSVEVLEFPDEEVVVGIPYLENLEDIIDSYSAPFTMONYLVRFLVD  
 BASE COUNT 686 a 700 c 766 g 542 t  
 ORIGIN  
 Query Match 92.0%; Score 18.4; DB 10; Length 2694;  
 Best Local Similarity 95.0%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TGGAGCGGCAGTGGCTGTG 20  
 ||||| ||||| ||||| |||||  
 Db 560 TGGAGCGCAGTGGCTGTG 579  
 RESULT 8  
 LOCUS AF157105 2892 bp mRNA linear ROD 25-NOV-1999  
 DEFINITION Mus musculus soluble secreted endopeptidase mRNA, complete cds.  
 ACCESSION AF157105  
 VERSION AF157105.1 GI:6467398  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2892)  
 AUTHORS Ikeda, K., Emoto, N., Raharjo, S.B., Nurhantari, Y., Saiki, K., Yokoyama, M. and Matsuo, M.  
 TITLE Molecular identification and characterization of novel membrane-bound metalloprotease, the soluble secreted form of which hydrolyzes a variety of vasoactive peptides  
 JOURNAL J. Biol. Chem. 274 (45), 32469-32477 (1999)  
 MEDLINE 20011457  
 PUBMED 10542292  
 REFERENCE 2 (bases 1 to 2892)  
 AUTHORS Ikeda, K., Emoto, N. and Matsuo, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JUN-1999) International Center for Medical Research, Kobe University School of Medicine, 7-5-1 Kusunoki, Chuo, Kobe 6500017, Japan  
 LOCATION/Qualifiers  
 source 1. .2892  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 CDS 281. .2578  
 /note="SEP; metalloprotease"  
 /codon\_start=1  
 /product="soluble secreted endopeptidase"  
 /protein\_id="AAF13152.1"  
 /db\_xref="GI:6467399"  
 /translation="MVERAGWCRKSPGFVEYGLMVLVLLLLGAIIVTLGVFYSIGKOL PLTSLHFSWDEITVVKRALRDSLSKSDICTTPSCVIAAARILENMQSRNCPENFY QYACGWLHRIHVIPETNSRYSVFDILRDELEVILKGVLEDSSTQHRPAVEKAKTLVRS CMNQSQVIEKRDSEPLLSVLMKVGWPMVAMDKWNETMGLKWELEROLAVLNSQFNRRVL IDLFINWDDQSSRRHVIYIDQPTLGMPSREYFQEDNNHKVKAYLEFMTSVATMLRK DONLSKESAMVREEMAEVLELEHLANATVQEKRDHVTALYHRMDLMELOERFLGK FNWTLFTQNLSSVEVLEFPDEEVVVGIPYLENLEDIIDSYSAPFTMONYLVRFLVD

RIGLSORFKEARDYRKALYGTVEEVRECVSVNMSAVGSLYIKRAFSDKS  
KSTVRELIEKIRSVFVNDLDELNMDESKKAKOECANNIREQIGYDPYILEDNNKHL  
DEEYSSLIFFYEDLIFENGLOKLNKNAQSLKLEKRVQDNLWIGAAYVNAFYSPNRN  
QIVFPAGILQPPFFSKDQPSLNFSGIGWIGHEITHGFDNGRNFKNMGLDWSN  
FSARHFOQSQCMYIYOYGNFWEADNONGVSTIGENIADNGVRQAYKAYLWLSA  
DGGKQORLPGLNLTYAQLFFINAYQVWCGSYRPEFAVQSIKTDVHSPLKRYVLGSLQN  
LPGFSAFHCPRGSPMHPMKRCRIW"

BASE COUNT 717 a 774 c 827 g 574 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 2892;  
Best Local Similarity 95.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGAGCGCGAGTGGCTGTG 20

||||| |||||||  
Db 885 TGGAGCGACAGTGGCTGTG 904

RESULT 9

AX033272

LOCUS AX033272 2925 bp DNA linear PAT 21-SEP-2000

DEFINITION Sequence 12 from Patent WO0047750.

ACCESSION AX033272

VERSION AX033272.1 GI:10280087

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2925)

AUTHORS Boileau, G. and Desgroselliers, L.

TITLE New metalloproteases of the neprilysin family

JOURNAL Patent: WO 0047750-A 12 17-AUG-2000;

BOILEAU GUY (CA); DESGROSSELLERS LUC (CA); UNIVERSITE DE MONTREAL

(CA)

FEATURES

source

1..2925 /organism="Mus musculus"

/db\_xref="taxon:10090"

332..2629

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC09977.1"

/db\_xref="GI:10280087"

/translation="MVERAGWCKKSPGFVEYGLMVLALLLGAIVTLGVFYSIGKQL

PLLTSLHFSDERTVVKRALRDSLSKSDICTPTSCVIAAARILENMDQSRPCENFY

QYACGGLRHVIVPTNRSYVFDILRDELEVLKGVLEDSQHRPAVEKATLYRS

CMNQSVIEKRDSEPLSVLKMVGWVPVAMKWNFTMGLKWELEKRLAVLNSVATMLRK

IDLFWDDQNSRHVYIDQPTLGMPSREYFQEDNNHKVRKAYLEFMTSVATMLRK

DONLSKESAMVREAEVLELETHLANATVPOEKRHDVTALYHRMDLMLQERFGLK

FNWTLFIQNLVSSVEVLPDEEVVYVIGIPYLENLEDIIDSYSARTMQLVWRLVLD

RIGLSORFKEARDYRKALYGTVEEVRECVSVNMSAVGSLYIKRAFSDKS

KSTVRELIEKIRSVFVNDLDELNMDESKKAKOECANNIREQIGYDPYILEDNNKHL

DEEYSSLIFFYEDLIFENGLOKLNKNAQSLKLEKRVQDNLWIGAAYVNAFYSPNRN

QIVFPAGILQPPFFSKDQPSLNFSGIGWIGHEITHGFDNGRNFKNMGLDWSN

FSARHFOQSQCMYIYOYGNFWEADNONGVSTIGENIADNGVRQAYKAYLWLSA

DGGKQORLPGLNLTYAQLFFINAYQVWCGSYRPEFAVQSIKTDVHSPLKRYVLGSLQN

LPGFSAFHCPRGSPMHPMKRCRIW"

BASE COUNT 710 a 797 c 836 g 582 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 2925;  
Best Local Similarity 95.0%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGAGCGCGAGTGGCTGTG 20

||||| |||||||  
Db 936 TGGAGCGACAGTGGCTGTG 955

RESULT 10

AF176569

LOCUS AF176569 2925 bp mRNA linear ROD 11-MAY-2000

DEFINITION Mus musculus neprilysin-like metalloproteinase 1 (N11) mRNA,

complete cds.

ACCESSION AF176569

VERSION AF176569.1 GI:7769082

KEYWORDS

SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2925)

AUTHORS Ghaddar, G., Ruchon, A.F., Carpentier, M., Marcinkiewicz, M.,

Seidah, N.G., Crine, P., Desgroselliers, L. and Boileau, G.

TITLE Molecular cloning and biochemical characterization of a new mouse

testis soluble-zinc-metalloproteinase of the neprilysin family

Biochem. J. 347 (Pt 2), 419-429 (2000)

21060448

10749671

REFERENCE 2 (bases 1 to 2925)

AUTHORS Ghaddar, G., Ruchon, A.F., Desgroselliers, L. and Boileau, G.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-1999) Biochemistry, University of Montreal, 2900.

Edouard Montpetit, Montreal, Que H3R 1J4, Canada

FEATURES

source

1..2925 /organism="Mus musculus"

/db\_xref="taxon:10090"

/tissue\_type="testis"

1..2925

/gene="N11"

332..2629

/gene="N11"

/note="N11; metallo-endopeptidase"

/codon\_start=1

/product="neprilysin-like metalloproteinase 1"

/protein\_id="AAF69247.1"

/db\_xref="GI:7769083"

/translation="MVERAGWCKKSPGFVEYGLMVLALLLGAIVTLGVFYSIGKQL

PLLTSLHFSDERTVVKRALRDSLSKSDICTPTSCVIAAARILENMDQSRPCENFY

QYACGGLRHVIVPTNRSYVFDILRDELEVLKGVLEDSQHRPAVEKATLYRS

CMNQSVIEKRDSEPLSVLKMVGWVPVAMKWNFTMGLKWELEKRLAVLNSVATMLRK

IDLFWDDQNSRHVYIDQPTLGMPSREYFQEDNNHKVRKAYLEFMTSVATMLRK

DONLSKESAMVREAEVLELETHLANATVPOEKRHDVTALYHRMDLMLQERFGLK

FNWTLFIQNLVSSVEVLPDEEVVYVIGIPYLENLEDIIDSYSARTMQLVWRLVLD

RIGLSORFKEARDYRKALYGTVEEVRECVSVNMSAVGSLYIKRAFSDKS

KSTVRELIEKIRSVFVNDLDELNMDESKKAKOECANNIREQIGYDPYILEDNNKHL

DEEYSSLIFFYEDLIFENGLOKLNKNAQSLKLEKRVQDNLWIGAAYVNAFYSPNRN

QIVFPAGILQPPFFSKDQPSLNFSGIGWIGHEITHGFDNGRNFKNMGLDWSN

FSARHFOQSQCMYIYOYGNFWEADNONGVSTIGENIADNGVRQAYKAYLWLSA

DGGKQORLPGLNLTYAQLFFINAYQVWCGSYRPEFAVQSIKTDVHSPLKRYVLGSLQN

LPGFSAFHCPRGSPMHPMKRCRIW"

BASE COUNT 710 a 797 c 836 g 582 t

ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 2925;

Best Local Similarity 95.0%; Pred. No. 1.1e+02;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGAGCGCGAGTGGCTGTG 20

||||| |||||||

Db 936 TGGAGCGACAGTGGCTGTG 955

RESULT 11

AL606829

LOCUS AL606829 203286 bp DNA linear ROD 05-APR-2002

DEFINITION Mouse DNA sequence from clone RP23-10M12 on chromosome 11, complete

sequence.

ACCESSION AL606829

VERSION AL606829.13 GI:20068543

KEYWORDS HTG.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1  
 AUTHORS Chapman, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 COMMENT On Apr 7, 2002 this sequence version replaced gi:18250804. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep). RP23-10M12 is from the RP23-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>. VECTOR: pBAC3.6.

## FEATURES

source Location/Qualifiers  
 1..203286  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="11"  
 /clone="RP23-10M12"  
 /clone\_lib="RPCI-23"

BASE COUNT 56507 a 42785 c 43119 g 60875 t  
 ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 203286;  
 Best Local Similarity 95.0%; Pred. No. 90;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
 ||||| ||||| ||||| |||||

Db 158012 TGGAGCGGCGAGTTGGCTGTG 158031

## RESULT 12

AL607032 208249 bp DNA linear HTG 24-JUL-2002  
 LOCUS Mus musculus chromosome 4 clone RP23-15L19, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION 2 unordered pieces.  
 ACCESSION AL607032  
 VERSION AL607032.15 GI:21955491  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 208249)

REFERENCE 1  
 AUTHORS Tracey, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JUL-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 COMMENT On Jul 25, 2002 this sequence version replaced gi:21627906.

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>

Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: BM15L19  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator; 11% of reads  
 Chemistry: Dye-terminator; Big Dye; 88% of reads  
 Consensus quality: 207834 bases at least Q40  
 Consensus quality: 207956 bases at least Q30  
 Consensus quality: 208060 bases at least Q20  
 Insert size: 208149; sum-of-contigs  
 Insert size: 219051; 2.3% error; agarose-fp  
 Quality coverage: 12.41x in Q20 bases; sum-of-contigs Quality coverage: 12.08x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* 1 137303: contig of 137303 bp in length  
 \* 137304 137403: gap of 100 bp  
 \* 137404 208249: contig of 70846 bp in length.

## FEATURES

source Location/Qualifiers  
 1..208249  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="4"  
 /clone="RP23-15L19"  
 /clone\_lib="RPCI-23"

## misc\_feature

1..137303  
 /note="assembly\_fragment:04071"

## misc\_feature

137404..208249  
 /note="assembly\_fragment:01838"  
 clone\_end:T7  
 vector\_side:right"

BASE COUNT 54098 a 53647 c 51328 g 49076 t 100 others  
 ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 208249;  
 Best Local Similarity 95.0%; Pred. No. 90;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
 ||||| ||||| ||||| |||||

Db 173552 TGGAGCGGCGAGTTGGCTGTG 173571

## RESULT 13

AC007116 25478 bp DNA linear HTG 30-JUL-1999  
 LOCUS Drosophila melanogaster chromosome 2 clone DS04467 (D447) map 60C6-60C8 strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION 5 unordered pieces.  
 ACCESSION AC007116  
 VERSION AC007116.2 GI:5630034  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 25478)

## REFERENCE

## AUTHORS

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,

```

Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
REFERENCE
2 (bases 1 to 25478)
Celnikier, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (19-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jul 28, 1999 this sequence version replaced gi:4454435.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. Pl library location:
47-51.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 587: contig of 687 bp in length
* 688 767: gap of unknown length
* 768 1326: contig of 559 bp in length
* 1327 1406: gap of unknown length
* 1407 2010: contig of 604 bp in length
* 2011 2090: gap of unknown length
* 2091 2817: contig of 727 bp in length
* 2818 2897: gap of unknown length
* 2898 25478: contig of 22581 bp in length.
Location/Qualifiers
1..25478
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="60C6-60C8"
/clone="DS04467 (D447)"
/clone_lib="Pl library, partial Sau3A in pAd10sacBII"
6515 a 6210 c 5900 g 5533 t 320 others
BASE COUNT
ORIGIN
Query Match 87.0%; Score 17.4; DB 2; Length 25478;
Best Local Similarity 94.7%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGAGCGGCGATGGTGTG 20
||||| ||||| ||||| |||||
Db 15038 GGAGCTGCAGTGGCTGTG 15056
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="60C6-60C8"
/clone="DS04467 (D447)"
/clone_lib="Pl library, partial Sau3A in pAd10sacBII"
82665 bp DNA linear HTG 17-FEB-2000
AC006938
Drosophila melanogaster chromosome 2 clone DS01630 (D506) map
60C7-60C8 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 9
unordered pieces.
AC006938
AC006938.4 GI:5633009
VERSION
KEYWORDS
HTG: HTGS.PHASE1.
SOURCE
Drosophila melanogaster.

```

```

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 82665)
Celnikier, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (02-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jul 28, 1999 this sequence version replaced gi:4936860.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgpf@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. Pl library location:
137-17.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 466: contig of 466 bp in length
* 467 546: gap of unknown length
* 547 1229: contig of 683 bp in length
* 1230 1309: gap of unknown length
* 1310 1884: contig of 575 bp in length
* 1885 1964: gap of unknown length
* 1965 2565: contig of 601 bp in length
* 2566 2645: gap of unknown length
* 2646 3310: contig of 665 bp in length
* 3311 3390: gap of unknown length
* 3391 3901: contig of 511 bp in length
* 3902 3982: gap of unknown length
* 3983 5536: contig of 1555 bp in length
* 5537 5616: gap of unknown length
* 5617 27193: contig of 21577 bp in length
* 27194 82665: contig of unknown length
* 27274 82665: contig of 55392 bp in length.
Location/Qualifiers
1..82665
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="60C7-60C8"
/clone="DS01630 (D506)"
/clone_lib="Pl library, partial Sau3A in pAd10sacBII"
21011 a 18780 c 19359 g 22875 t 640 others
BASE COUNT
ORIGIN
Query Match 87.0%; Score 17.4; DB 2; Length 82665;

```

Best Local Similarity 94.7%; Pred. No. 3.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTGTG 20  
||||| ||||||| |||||||  
Db 23273 GGAGCTGCAGTTGGCTGTG 23255

## RESULT 15

AC096785\_1

WPCOMMENT

Sequence split into 4 fragments LOCUS AC096785 Accession AC096785

Fragment Name	Begin	End
AC096785_0	1	110000
AC096785_1	100001	210000
AC096785_2	200001	310000
AC096785_3	300001	402740

Continuation (2 of 4) of AC096785 from base 100001 (AC096785 Mus musculus chromosome 18)

Query Match 87.0%; Score 17.4; DB 2; Length 110000;  
Best Local Similarity 94.7%; Pred. No. 3.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGT 19  
||||| ||||||| |||||||  
Db 98215 TGGAGCAGCAGTTGGCTGT 98233

Search completed: July 8, 2003, 03:34:17  
Job time : 224.098 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 125.659 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-5

Perfect score: 20

Sequence: 1 tggagcggcagtgctgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

```

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Score	Match	Length	ID	Description
1	20	100.0	20	AZ228811	Rat membrane metal
2	20	100.0	2286	AAD28130	Soluble secreted e
3	20	100.0	2765	AZ228810	Rat membrane metal
4	18.4	92.0	2925	AAA63763	cDNA encoding neut
5	16.8	84.0	3439	ABL03141	Drosophila melanog
6	16.8	84.0	4308	ABL14154	Drosophila melanog
7	16.8	84.0	7620	ABL03159	Drosophila melanog
8	16.8	84.0	9695	ABL03140	Drosophila melanog
9	16.8	84.0	9820	ABL03158	Drosophila melanog

10	16.4	82.0	457	24	ABL90759	Human polynucleoti
11	16.4	82.0	1004	23	ABL09885	Drosophila melanog
12	16.4	82.0	3178	23	ABL09884	Drosophila melanog
13	16.4	82.0	3230	22	AAH15715	Human cDNA sequenc
14	16.4	82.0	4179	22	AAH98647	Hamster EST-derive
15	16.4	82.0	4180	22	AAH98669	Hamster EST-derive
16	16.4	82.0	6779	23	ABLI0856	Drosophila melanog
17	16.4	82.0	9234	22	AAK89255	Human digestive sy
18	16.4	82.0	22452	22	AAK27827	DNA encoding novel
19	16.4	82.0	22452	22	AAK27829	DNA encoding novel
20	15.8	79.0	395	22	ABA08497	Human ionotropic G
21	15.8	79.0	1195	16	AAQ79154	A. aculeatus pecti
22	15.8	79.0	1419	24	ABQ68965	Listeria monocytog
23	15.8	79.0	2936	21	AAA09315	Human cancer assoc
24	15.8	79.0	2936	24	ABK13306	DNA encoding renal
25	15.8	79.0	3163	20	AAK60266	Nucleic acid seque
26	15.8	79.0	3714	24	AAK33652	Human TRICH-7 CDNA
27	15.8	79.0	4197	24	AAK36472	Human N-methyl-D-a
28	15.8	79.0	4896	24	ABQ70960	Listeria monocytog
29	15.8	79.0	5746	22	AAAL03142	Human reproductive
30	15.8	79.0	7789	24	AAI16904	cDNA encoding huma
31	15.8	79.0	12683	22	AAI98939	Human excretory re
32	15.8	79.0	12683	22	AAI63289	Human kidney relat
33	15.8	79.0	12744	22	AAI98938	Human excretory re
34	15.8	79.0	12744	22	AAI63288	Human kidney relat
35	15.8	79.0	74586	24	AAI16905	Genomic DNA encodi
36	15.8	79.0	1230025	20	AAK91990	Nucleotide sequenc
37	15.4	77.0	5274	23	ABL15583	Drosophila melanog
38	15.4	77.0	8785	23	ABL15582	Drosophila melanog
39	15.4	77.0	11293	22	ABA17918	Human nervous syst
40	15.4	77.0	22107	22	ABA17921	Human nervous syst
41	15.4	77.0	22107	22	AAK79542	Human immune/haema
42	15.4	77.0	22109	22	ABA17922	Human nervous syst
43	15.4	77.0	22109	22	AAK79543	Human immune/haema
44	15.4	77.0	22111	22	ABA17919	Human nervous syst
45	15.4	77.0	22111	22	AAK79540	Human immune/haema

#### ALIGNMENTS

#### RESULT 1

AZ228811 standard; DNA; 20 BP.

AC AZ228811;

DT 01-FEB-2000 (first entry)

XX Rat membrane metalloprotease NEPII gene probe #1.

DE Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.

OS Synthetic.

OS Rattus rattus.

XX FR2777291-A1.

XX 15-OCT-1999.

PF 08-APR-1998; 98FR-0004389.

XX 08-APR-1998; 98FR-0004389.

PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;

PI Schwartz JC;

XX WPI; 1999-593429/51.

XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease  
XX  
PS Claim 3; Page 20; 29pp; French.  
XX  
CC Sequences AAZ28811-228827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysine II (NEPII) gene  
CC (AAZ28810). NEPII is involved in (in)activation of neuronal and hormonal  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
CC disorders.  
XX  
SQ Sequence 20 BP; 2 A; 3 C; 10 G; 5 T; 0 other;  
Query Match 100.0%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TGGAGCGGCAGTTGGCTGTG 20  
Db 1 TGGAGCGGCAGTTGGCTGTG 20  
  
RESULT 2  
AAZ28130  
ID AAZ28130 standard; DNA; 2286 BP.  
XX  
AC AAZ28130;  
XX  
DT 07-MAY-2002 (first entry)  
XX  
DE Soluble secreted endopeptidase (SEP) consensus DNA.  
XX  
KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
KW FSAD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
XX  
OS Homo sapiens.  
OS Mus sp.  
OS Rattus sp.  
XX  
FH Key Location/Qualifiers  
FT misc\_feature 1664..2286  
FT /tag= a  
FT /note= "Encodes catalytic domain"  
XX  
PN WO200206492-A1.  
XX  
PD 24-JAN-2002.  
XX  
PF 16-JUL-2001; 2001WO-IB01263.  
XX  
PR 14-JUL-2000; 2000GB-0017387.  
XX  
PA (PFIZ ) PFIZER LTD.  
PA (PFIZ ) PFIZER INC.  
XX  
PI Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
XX  
DR WPI; 2002-155042/20.  
XX  
XX An isolated and/or purified nucleic acid encoding a human soluble  
PT secreted endopeptidase which is useful for treating sexual dysfunction,  
PT for e.g. male erectile dysfunction, or female sexual dysfunction such as  
PT female sexual arousal disorder -  
XX

PS Disclosure; Fig 6; 167pp; English.  
XX  
CC The invention relates to an isolated and/or purified nucleic acid  
CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
CC antibody and the compound which inhibits or selectively inhibits the  
CC human SEP protein are useful in the manufacture of a medicament for  
CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
CC also useful for treating the above disorders and other disorders such  
CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
CC hypoaffective sexual desire disorder. The present sequence is SEP consensus  
CC DNA sequence found in human, mouse and rat.  
XX  
SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
Query Match 100.0%; Score 20; DB 24; Length 2286;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 TGGAGCGGCAGTTGGCTGTG 20  
Db 617 TGGAGCGGCAGTTGGCTGTG 636  
  
RESULT 3  
AAZ28810  
ID AAZ28810 standard; cDNA; 2765 BP.  
XX  
AC AAZ28810;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Rat membrane metalloprotease NEPII gene.  
XX  
KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
XX  
OS Rattus rattus.  
XX  
PN PR2777291-A1.  
XX  
PD 15-OCT-1999.  
XX  
PF 08-APR-1998; 98FR-0004389.  
XX  
PR 08-APR-1998; 98FR-0004389.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Guilmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
XX  
DR WPI; 1999-593429/51.  
DR P-PSDB; AAY44177.  
XX  
PT New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease -  
XX  
PS Claim 2; Page 12-16; 29pp; French.  
XX  
CC This sequence represents the gene for the rat membrane metalloprotease  
CC designated neprilysine II (NEPII), which is involved in (in)activation  
CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
CC for specific substrates (used to detect NEPII in cells and tissues) or  
CC inhibitors, which can also be used to detect NEPII or for treatment of  
CC disorders related to peptidergic signalling in which NEPII is involved,  
CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
CC endocrine disorders.  
XX

```
XX SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;
Query Match 100.0%; Score 20; DB 20; Length 2765;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20
   ||||| ||||| ||||| |||||
Db 741 TGGAGCGGCAGTTGGCTGTG 760

RESULT 4
AAA63763
ID AAA63763 standard; cDNA; 2925 BP.
XX AC AAA63763;
XX 04-DEC-2000 (first entry)
XX cDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.
KW Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;
KW NEP-like enzyme; protein production; protein secretion;
KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;
KW fertility; bone disease; abnormal phosphate metabolism; ss.
XX OS Mus sp.
XX Key Location/Qualifiers
FH 332..2629
FT /*tag= a
FT /product= "neutral endopeptidase metalloproteinase-like
FT enzyme NL-1"
XX WO200047750-A2.
XX 17-AUG-2000.
XX 11-FEB-2000; 2000WO-CA00147.
XX 11-FEB-1999; 99CA-2260376.
XX (UYMO-) UNIV MONTREAL.
XX Desgroseillers L, Boileau G;
XX WPI; 2000-549148/50.
XX P-PSDB; AAB08130.
XX Novel neutral endopeptidase-like metalloproteinase polypeptides and
XX polynucleotides, used to screen for related sequences and enzyme
XX inhibitors, used for the treatment of NL-3 related bone disorders -
XX Disclosure; Fig 3; 59pp; English.
XX The present sequence encodes a murine neutral endopeptidase
XX metalloproteinase-like enzyme, designated NL-1. The specification
XX also describes NL-2 and NL-3. The NL enzymes are used to test for
XX specific inhibitors. The N-terminal region of the enzymes can be used
XX to promote production and secretion of foreign proteins and active
XX biopeptides, using chimeric constructs containing the foreign protein
XX downstream from and in phase with the N-terminal region. The NL enzymes
XX are have been localised to the brain, and may be useful in the
XX treatment of neurological diseases such as Alzheimer's disease, pain,
XX and psychiatric disorders. NL enzymes have also been localised to the
XX testis and ovaries, and may be used to control fertility. They have
XX also been localised to bones, and may be used to treat bone diseases,
XX and abnormal phosphate metabolisms related to improper peptide
XX processing by the NL-3 enzyme.
XX SQ Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;
```

```
Query Match 92.0%; Score 18.4; DB 21; Length 2925;
Best Local Similarity 95.0%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20
   ||||| ||||| ||||| |||||
Db 936 TGGAGCGGCAGTTGGCTGTG 955

RESULT 5
ABL03141
ID ABL03141 standard; cDNA; 3439 BP.
XX AC ABL03141;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster, expressed polynucleotide SEQ ID NO 3905.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX P-PSDB; ABB59038.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Claim 1; SEQ ID NO 3905; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3439 BP; 899 A; 923 C; 979 G; 638 T; 0 other;

Query Match 84.0%; Score 16.8; DB 23; Length 3439;
Best Local Similarity 90.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20
   ||||| ||||| ||||| |||||
Db 1309 TGGAGCGGCAGTTAGTGTG 1328

RESULT 6
ABL14154/c
ID ABL14154 standard; cDNA; 4308 BP.
XX
```

AC ABL14154;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36944.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB70051.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 36944; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4308 BP; 1096 A; 1031 C; 959 G; 1222 T; 0 other;  
XX  
Query Match 84.0%; Score 16.8; DB 23; Length 4308;  
Best Local Similarity 90.0%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
QY 1 TGGAGCGGCAGTTGGCTGTG 20  
DB 1741 TGGAGCGGCAGTTGGCTGTG 1722  
XX  
RESULT 7  
ABL03159  
ID ABL03159 standard; cDNA; 7620 BP.  
XX  
AC ABL03159;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 3959.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB59037.

PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36944.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB59056.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1; SEQ ID NO 3959; 2lpp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 7620 BP; 2124 A; 2020 C; 2109 G; 1367 T; 0 other;  
XX  
Query Match 84.0%; Score 16.8; DB 23; Length 7620;  
Best Local Similarity 90.0%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
XX  
QY 1 TGGAGCGGCAGTTGGCTGTG 20  
DB 5490 TGGAGCGGCAGTTAGCTGTG 5509  
XX  
RESULT 8  
ABL03140  
ID ABL03140 standard; cDNA; 9695 BP.  
XX  
AC ABL03140;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 3902.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB59037.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
XX  
PS Claim 1; SEQ ID NO 3902; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 9695 BP; 2668 A; 2438 C; 2581 G; 2008 T; 0 other;  
  
Query Match 84.0%; Score 16.8; DB 23; Length 9695;  
Best Local Similarity 90.0%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 TGGAGCGCGACTTGGCTGTG 20  
||||| |||||| ||||||  
Db 6436 TGGAGCAGCAGTTAGCTGTG 6455  
  
RESULT 9  
ABL03158  
ID ABL03158 standard; cDNA; 9820 BP.  
XX  
AC ABL03158;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 3956.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmacological; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX P-PSDB; ABB59055.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
XX  
PS Claim 1; SEQ ID NO 3956; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 9820 BP; 2698 A; 2466 C; 2628 G; 2028 T; 0 other;  
  
Query Match 84.0%; Score 16.8; DB 23; Length 9820;  
Best Local Similarity 90.0%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 TGGAGCGCGACTTGGCTGTG 20  
||||| |||||| ||||||  
Db 6561 TGGAGCAGCAGTTAGCTGTG 6580  
  
RESULT 10  
ABL90759  
ID ABL90759 standard; cDNA; 457 BP.  
XX  
XX ABL90759;  
XX  
XX 24-MAY-2002 (first entry)  
XX  
XX Human polynucleotide SEQ ID NO 1321.  
DE  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; anti-ulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200190304-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 18-MAY-2001; 2001WO-US16450.  
XX  
XX 19-MAY-2000; 2000US-205515P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Birse CE, Rosen CA;  
XX  
XX WPI; 2002-122018/16.  
XX  
XX P-PSDB; ABB90350.  
XX  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders  
XX  
XX Claim 4; SEQ ID NO 1321; 2081pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins  
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
XX  
XX Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 457 BP; 73 A; 163 C; 119 G; 101 T; 1 other;

Query Match 82.0%; Score 16.4; DB 24; Length 457;  
Best Local Similarity 94.4%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GAGCGGCAGTGGCTGTG 20

Db 247 GAGCGGCAGTGGCTGTG 264

## RESULT 11

ABL09885/C  
ID ABL09885 standard; cDNA; 1004 BP.

XX AC

XX ABL09885;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 24137.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS

XX WO200171042-A2.

PN

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB65782.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 24137; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1004 BP; 304 A; 258 C; 265 G; 177 T; 0 other;

SQ

Query Match 82.0%; Score 16.4; DB 23; Length 1004;

Best Local Similarity 94.4%; Pred. No. 2.4e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTGGCTGTG 18

Db 950 TGGAGCGGCAGTGGCTGTG 933

## RESULT 12

ABL09884/C  
ID ABL09884 standard; cDNA; 3178 BP.

XX AC

XX ABL09884;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 24134.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS

XX WO200171042-A2.

PN

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB65781.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 24134; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX sequences (ABL01840-ABL16175) and the encoded proteins

XX (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3178 BP; 952 A; 702 C; 725 G; 799 T; 0 other;

SQ

Query Match 82.0%; Score 16.4; DB 23; Length 3178;

Best Local Similarity 94.4%; Pred. No. 2.7e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTGGCTGTG 18

Db 2124 TGGAGCGGCAGTGGCTGTG 2107

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXX

OS Homo sapiens.  
 XX EP1074617-A2.  
 XX PD 07-FEB-2001.  
 XX PF 28-JUL-2000; 2000EP-0116126.  
 XX PR 29-JUL-1999; 99JP-0248036.  
 XX PR 27-AUG-1999; 99JP-030253.  
 XX PR 11-JAN-2000; 2000JP-0118776.  
 XX PR 02-MAY-2000; 2000JP-0183767.  
 XX PR 09-JUN-2000; 2000JP-0241899.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX DR Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs.  
 XX PS Claim 8; SEQ ID 14100; 2537pp + CD ROM; English.  
 XX CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX SQ Sequence 3230 BP; 579 A; 1041 C; 977 G; 633 T; 0 other;  
 Query Match 82.0%; Score 16.4; DB 22; Length 3230;  
 Best Local Similarity 94.4%; Pred. No. 2.7e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 GAGCGGCAGTGGCTGTG 20  
 Db 675 GAGCGGCAGTGGCTGTG 692  
 RESULT 14  
 AAH98647  
 ID AAH98647 standard; cDNA; 4179 BP.  
 AC AAH98647;  
 XX 12-OCT-2001 (first entry)  
 DT Hamster EST-derived coding sequence SEQ ID NO: 504.  
 DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 XX

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition; ss.  
 XX Cricetulus griseus.  
 XX OS WO200154477-A2.  
 XX PN 02-AUG-2001.  
 XX PD 25-JAN-2001; 2001WO-US02687.  
 XX PF 25-JAN-2000; 2000US-0491404.  
 XX PR 17-JUL-2000; 2000US-0617746.  
 XX PR 03-AUG-2000; 2000US-0631451.  
 XX PR 15-SEP-2000; 2000US-0663870.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Drmanac RA, Zhang J, Werhman T;  
 XX WPI; 2001-476164/51.  
 XX DR P-PSDB; AAM23988.  
 XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use.  
 XX PS Claim 1; Page 529-531; 1275pp; English.  
 XX CC The present invention provides the protein and coding sequences of novel  
 CC proteins from a variety of organisms, including human, dog, cat, horse,  
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 CC from the organism of interest. They can be used in diagnostics,  
 CC forensics, gene mapping, identification of mutations, to assess  
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA  
 CC of the invention.  
 XX SQ Sequence 4179 BP; 767 A; 1323 C; 1224 G; 865 T; 0 other;  
 Query Match 82.0%; Score 16.4; DB 22; Length 4179;  
 Best Local Similarity 94.4%; Pred. No. 2.8e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 GAGCGGCAGTGGCTGTG-20  
 Db 1609 GAGCGGCAGTGGCTGTG 1626  
 RESULT 15  
 AAH98669  
 ID AAH98669 standard; cDNA; 4180 BP.  
 AC AAH98669;  
 XX 12-OCT-2001 (first entry)  
 DT Hamster EST-derived coding sequence SEQ ID NO: 526.  
 DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition; ss.  
 XX Cricetulus griseus.  
 XX OS WO200154477-A2.  
 XX PN 02-AUG-2001.  
 XX PD 25-JAN-2001; 2001WO-US02687.  
 XX PF 25-JAN-2000; 2001WO-US02687.



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 27.6098 Seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-5

Perfect score: 20

Sequence: 1 tggagcgagctgtgctgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A-COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5B-COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A-COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B-COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS-COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.8	79.0	541	4	US-09-465-558-31
C 2	15.8	79.0	1195	1	US-08-535-230A-1
C 3	15.8	79.0	2935	4	US-09-166-350-6
C 4	15.2	76.0	479	4	US-09-465-558-25
C 5	15.2	76.0	536	4	US-09-465-558-63
C 6	15.2	76.0	1079	1	US-08-270-583-1
C 7	15.2	76.0	1079	1	US-08-783-889A-1
C 8	15.2	76.0	1102	3	US-09-067-089-1
C 9	15.2	76.0	1297	6	5187077-23
C 10	15.2	76.0	1297	6	5427925-21
C 11	15.2	76.0	1468	4	US-09-465-558-69
C 12	15.2	76.0	1554	3	US-09-109-205-4
C 13	15.2	76.0	1554	4	US-09-443-184-46
C 14	15.2	76.0	1558	4	US-09-123-030-7
C 15	15.2	76.0	2265	1	US-08-242-689-1
C 16	15.2	76.0	2265	4	US-08-468-583-1
C 17	15.2	76.0	3218	4	US-09-221-017B-255
C 18	15.2	76.0	43360	4	US-09-453-702B-206
C 19	15.2	76.0	45325	4	US-09-453-702B-261
C 20	15.2	76.0	4403765	4	US-09-103-840A-2
C 21	15.2	76.0	4411529	4	US-09-103-840A-1
C 22	14.8	74.0	217	1	US-08-791-495-1
C 23	14.8	74.0	345	4	US-09-199-637A-256
C 24	14.8	74.0	384	4	US-09-199-637A-254
C 25	14.8	74.0	592	4	US-08-896-164-80
C 26	14.8	74.0	752	1	US-08-725-182C-1
C 27	14.8	74.0	752	3	US-09-013-150-1

Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 6, Appli  
Sequence 8, Appli  
Sequence 3, Appli  
Sequence 192, App  
Sequence 191, App  
Sequence 190, App  
Sequence 189, App  
Sequence 252, App  
Sequence 102, App  
Sequence 128, App  
Sequence 64, Appli  
Sequence 1, Appli  
Sequence 2, Appli

#### ALIGNMENTS

RESULT 1  
US-09-465-558-31/c  
; Sequence 31, Application US/09465558  
; Patent No. 6436657  
; GENERAL INFORMATION:  
; APPLICANT: Morakinyo, Layo O.  
; APPLICANT: Orozco, Jr, Emil M.  
; TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
; FILE REFERENCE: B1322 US NA  
; CURRENT APPLICATION NUMBER: US/09465,558  
; CURRENT FILING DATE: 1999-12-17  
; EARLIER APPLICATION NUMBER: 60/112,734  
; EARLIER FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 31  
; LENGTH: 541  
; TYPE: DNA  
; ORGANISM: Triticum aestivum  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (260)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (369)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (374)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (447)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (455)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (465)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (468)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (477)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (500)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (507)  
; FEATURE:

C 28 14.8 74.0 752 4 US-09-359-503-1  
C 29 14.8 74.0 752 4 US-09-062-422-1  
C 30 14.8 74.0 752 4 US-09-396-184-1  
C 31 14.8 74.0 752 4 US-08-937-263B-1  
C 32 14.8 74.0 755 1 US-08-791-495-6  
C 33 14.8 74.0 755 1 US-08-791-495-8  
C 34 14.8 74.0 1002 1 US-08-791-495-4  
C 35 14.8 74.0 1002 2 US-08-359-850-3  
C 36 14.8 74.0 1653 4 US-09-199-637A-192  
C 37 14.8 74.0 1827 4 US-09-199-637A-191  
C 38 14.8 74.0 1920 4 US-09-199-637A-190  
C 39 14.8 74.0 1932 4 US-09-199-637A-189  
C 40 14.8 74.0 1938 4 US-09-199-637A-252  
C 41 14.8 74.0 2101 4 US-09-199-637A-102  
C 42 14.8 74.0 13440 4 US-08-961-527-128  
C 43 14.8 74.0 31728 4 US-09-453-702B-64  
C 44 14.8 74.0 42235 4 US-09-199-637A-1  
C 45 14.8 74.0 4403765 4 US-09-103-840A-2

NAME/KEY: unsure  
LOCATION: (517)...(518)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (520)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (530)  
US-09-465-558-31

Query Match 79.0%; Score 15.8; DB 4; Length 541;  
Best Local Similarity 85.0%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
DB 263 TGGNGCGGCAGTTGACGGTG 244

## RESULT 2

US-08-535-230A-1/c  
Sequence 1, Application US/08535230A  
Patent No. 5707847

## GENERAL INFORMATION:

APPLICANT: Christgau, Stephan  
APPLICANT: Kofod, Lene Kofod  
APPLICANT: Andersen, Lene No. 5707847boe  
APPLICANT: Kauppinen, Sakari  
APPLICANT: Heldt-Hansen, Hans Peter  
APPLICANT: Budolfson, Gitte  
APPLICANT: Dalboge, Henrik  
TITLE OF INVENTION: An Enzyme Exhibiting Pectin  
TITLE OF INVENTION: Methylsterase  
NUMBER OF SEQUENCES: 18

## CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5707847o No. 5707847disk of No. 5707847th America, Inc.  
STREET: 405 Lexington Avenue  
CITY: New York  
STATE: NY

## COUNTRY: USA

## ZIP: 10174-6401

## COMPUTER READABLE FORM:

## MEDIUM TYPE: Diskette

## COMPUTER: IBM Compatible

## OPERATING SYSTEM: DOS

## SOFTWARE: FastSeq for Windows Version 2.0

## CURRENT APPLICATION DATA:

## APPLICATION NUMBER: US/08/535,230A

## FILING DATE: 02-NOV-1995

## CLASSIFICATION:

## ATTORNEY/AGENT INFORMATION:

## NAME: Valeta, Gregg A

## REGISTRATION NUMBER: 35,127

## REFERENCE/DOCKET NUMBER: 3667,204-US

## TELECOMMUNICATION INFORMATION:

## TELEPHONE: 212-867-0123

## TELEFAX: 212-878-9655

## TELEX:

## INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

## LENGTH: 1195 base pairs

## TYPE: nucleic acid

## STRANDEDNESS: single

## TOPOLOGY: linear

## US-08-535-230A-1

Query Match 79.0%; Score 15.8; DB 1; Length 1195;  
Best Local Similarity 89.5%; Pred. No. 75;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTGTG 20  
||||| ||||| ||||| |||||  
DB 679 GGAGCGCGCGTTGGCGGTG 661

## RESULT 3

US-09-166-350-6  
Sequence 6, Application US/09166350A  
Patent No. 6446663

## GENERAL INFORMATION:

APPLICANT: Scanlan, Matthew  
APPLICANT: Chen, Yao  
APPLICANT: Stockert, Elisabeth  
APPLICANT: Old, Lloyd  
APPLICANT: Jager, Elke  
APPLICANT: Knuth, Alex  
TITLE OF INVENTION: Renal Cancer Associated Antigens and  
FILE REFERENCE: L0461/7051  
CURRENT APPLICATION NUMBER: US/09/166,350A  
CURRENT FILING DATE: 1998-10-05  
EARLIER APPLICATION NUMBER: US 09/166,350  
EARLIER FILING DATE: 1998-10-05  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6

## LENGTH: 2936

## TYPE: DNA

## ORGANISM: Homo sapiens

## US-09-166-350-6

Query Match 79.0%; Score 15.8; DB 4; Length 2936;  
Best Local Similarity 89.5%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||

DB 1058 GGTGCGGCAGTTGGCTGAG 1076

## RESULT 4

US-09-465-558-25/C  
Sequence 25, Application US/09465558  
Patent No. 6436657

## GENERAL INFORMATION:

APPLICANT: Morakinyo, Layo O.  
APPLICANT: Orozco Jr, Emil M.  
TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
FILE REFERENCE: BB1322 US NA  
CURRENT APPLICATION NUMBER: US/09/465,558  
CURRENT FILING DATE: 1999-12-17  
EARLIER APPLICATION NUMBER: 60/112,734  
EARLIER FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 25

## LENGTH: 479

## TYPE: DNA

## ORGANISM: zea mays

## FEATURE:

## NAME/KEY: unsure

## LOCATION: (391)

## FEATURE:

## NAME/KEY: unsure

## LOCATION: (399)

## FEATURE:

## NAME/KEY: unsure

## LOCATION: (429)

## FEATURE:

## NAME/KEY: unsure

## LOCATION: (436)

## US-09-465-558-25

Query Match 76.0%; Score 15.2; DB 4; Length 479;  
Best Local Similarity 85.0%; Pred. No. 1,3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
||| ||||||||| |||  
Db 284 TGGCGCGGCGAGTTGACGGTG 265

## RESULT 5

US-09-465-558-63/C  
; Sequence 63, Application US/09465558  
; Patent No. 6436657  
; GENERAL INFORMATION:  
; APPLICANT: Morakinyo, Layo O.  
; APPLICANT: Orozco Jr, Emil M.  
; TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
; FILE REFERENCE: BB1322 US NA  
; CURRENT APPLICATION NUMBER: US/09/465,558  
; EARLIER FILING DATE: 1999-12-17  
; EARLIER APPLICATION NUMBER: 60/112,734  
; EARLIER FILING DATE: 1998-12-18  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 63  
; LENGTH: 536  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (329)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (378)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (390)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (406)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (410)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (419)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (425)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (442)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (457)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (465)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (470)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (477)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (479)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (485)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (494)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (508)

; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (520)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (525)  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (531)  
; US-09-465-558-63

Query Match 76.0%; Score 15.2; DB 4; Length 536;

Best Local Similarity 85.0%; Pred. No. 1.3e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
||| ||||||||| |||  
Db 289 TGGCGCGGCGAGTTGACGGTG 270

## RESULT 6

US-08-270-583-1  
; Sequence 1, Application US/08270583  
; Patent No. 5629193  
; GENERAL INFORMATION:  
; APPLICANT: HUDSON, ET AL.  
; TITLE OF INVENTION: Serum Paraoxonase  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/270,583  
; FILING DATE: CONCURRENTLY  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 325800-131  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1079 BASE PAIRS  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: CDNA  
; US-08-270-583-1

Query Match 76.0%; Score 15.2; DB 1; Length 1079;

Best Local Similarity 85.0%; Pred. No. 1.4e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
||| ||||||| |||||||  
Db 5 TGGCGCGGCGAGTTGGCTGTG 24

RESULT 7  
US-08-783-889A-1  
; Sequence 1, Application US/08783889A  
; Patent No. 5792639  
; GENERAL INFORMATION:  
; APPLICANT: PETER L. HUDSON, ET AL.  
; TITLE OF INVENTION: Serum Paraoxonase  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/783,889A  
; FILING DATE: 16 JAN 1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/270,583  
; FILING DATE: 5 JULY 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MULLINS, J.G.  
; REGISTRATION NUMBER: 33,073  
; REFERENCE/DOCKET NUMBER: 325800-594 (PFI24D1)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1079 BASE PAIRS  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: cDNA  
US-08-783-889A-1

Query Match 76.0%; Score 15.2; DB 1; Length 1079;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGCGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
DB 5 TGGGCGGCTGTGGCTGTG 24

RESULT 8  
US-09-067-089-1  
; Sequence 1, Application US/09067089A  
; Patent No. 6140093  
; GENERAL INFORMATION:  
; APPLICANT: Hudson, Peter L.  
; APPLICANT: He, Wei W.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Serum Paraoxonase  
; FILE REFERENCE: PFI24D2  
; CURRENT APPLICATION NUMBER: US/09/067,089A  
; CURRENT FILING DATE: 1998-04-27  
; EARLIER APPLICATION NUMBER: 08/783,889  
; EARLIER FILING DATE: 1997-01-16  
; EARLIER APPLICATION NUMBER: 08/270,583  
; EARLIER FILING DATE: 1994-07-05  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1102

; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (38)..(1102)  
US-09-067-089-1

Query Match 76.0%; Score 15.2; DB 3; Length 1102;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGCGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
DB 39 TGGGCGGCTGTGGCTGTG 58

RESULT 9  
5187077-23  
; Patent No. 5187077  
; APPLICANT: GEARING, DAVID P.; GOUGH, NICHOLAS M.; HILTON,  
; DOUGLAS J.; KING, JULIE A.; METCALF, DONALD; NICE, EDOUARD C.  
; NICOLA, NICOS A.; SIMPSON, RICHARD J.; WILLSON, TRACY A.  
; TITLE OF INVENTION: LEUKEMIA INHIBITORY FACTOR  
; NUMBER OF SEQUENCES: 41  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/294,514  
; FILING DATE: 09-DEC-1988  
; SEQ ID NO:23:  
; LENGTH: 1297  
5187077-23

Query Match 76.0%; Score 15.2; DB 6; Length 1297;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGCGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
DB 504 TTGAGAGGCAGTGGGCTGTG 523

RESULT 10  
5427925-21  
; Patent No. 5427925  
; APPLICANT: GEARING, DAVID P.; GOUGH, NICHOLAS M.; HILTON,  
; DOUGLAS J.; KING, JULIE A.; METCALF, DONALD; NICE, EDOUARD C.  
; NICOLA, NICOS A.; SIMPSON, RICHARD J.; WILLSON, TRACY A.  
; TITLE OF INVENTION: RECOMBINANT METHOD FOR MAKING  
; LEUKEMIA INHIBITOR FACTOR  
; NUMBER OF SEQUENCES: 38  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/58,979  
; FILING DATE: 06-MAY-1993  
; APPLICATION NUMBER: 948,614  
; FILING DATE: 22-SEP-1992  
; APPLICATION NUMBER: 667,159  
; FILING DATE: 11-MAR-1991  
; SEQ ID NO:21:  
; LENGTH: 1297  
5427925-21

Query Match 76.0%; Score 15.2; DB 6; Length 1297;  
Best Local Similarity 85.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGCGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
DB 504 TTGAGAGGCAGTGGGCTGTG 523

RESULT 11  
US-09-465-558-69/c  
; Sequence 69, Application US/09465558  
; Patent No. 6436657

GENERAL INFORMATION:  
APPLICANT: Morakinyo, Layo O.  
APPLICANT: Orozco Jr., Emil M.  
TITLE OF INVENTION: TETRAHYDROFOLATE METABOLIC ENZYMES  
FILE REFERENCE: BB1322 US NA  
CURRENT APPLICATION NUMBER: US/09/465,558  
CURRENT FILING DATE: 1999-12-17  
EARLIER APPLICATION NUMBER: 60/112,734  
EARLIER FILING DATE: 1998-12-18  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 69  
LENGTH: 1468  
TYPE: DNA  
ORGANISM: Triticum aestivum  
US-09-465-558-69

Query Match 76.0%; Score 15.2; DB 4; Length 1468;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
Db 270 TGGCGCGGCAGTTGACGGTG 251

RESULT 12  
US-09-109-205-4  
Sequence 4, Application US/09109205  
Patent No. 6057140.  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Gorgone, Gina  
APPLICANT: Corley, Neil C.  
APPLICANT: Baughn, Mariah R.  
APPLICANT: Yue, Henry  
TITLE OF INVENTION: HUMAN SCAD FAMILY MOLECULES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/109,205  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerrone, Michael C  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0542 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-855-0572  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1554 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:

LIBRARY: PROSTUT12  
CLONE: 1810320  
US-09-109-205-4

Query Match 76.0%; Score 15.2; DB 3; Length 1554;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
Db 657 TGACGGCGCAGTTGGCTGTG 676

RESULT 13  
US-09-443-184-46  
Sequence 46, Application US/09443184A  
Patent No. 6372431  
GENERAL INFORMATION:  
APPLICANT: Cunningham, Mary Jane  
APPLICANT: Zweiger, Gary  
APPLICANT: Kaser, Matthew R.  
APPLICANT: Panzer, Scott  
APPLICANT: Seilhammer, Jeffrey J.  
APPLICANT: Yue, Henry  
APPLICANT: Baughn, Mariah  
APPLICANT: Azimzal, Yalda  
APPLICANT: Lal, Preeti  
TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS  
FILE REFERENCE: PG-0007 US  
CURRENT APPLICATION NUMBER: US/09/443,184A  
CURRENT FILING DATE: 1999-11-19  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: PERL Program  
SEQ ID NO 46  
LENGTH: 1554  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. 6372431 1810320CB1  
US-09-443-184-46

Query Match 76.0%; Score 15.2; DB 4; Length 1554;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
||| ||||| ||||| |||||  
Db 657 TGACGGCGCAGTTGGCTGTG 676

RESULT 14  
US-09-123-030-7/c  
Sequence 7, Application US/09123030  
Patent No. 6365337  
GENERAL INFORMATION:  
APPLICANT: Letts, Verity A.  
APPLICANT: Frankel, Wayne N.  
APPLICANT: Campbell, Kevin P.  
APPLICANT: Felix, Ricardo  
APPLICANT: Biddlecome, Gloria  
TITLE OF INVENTION: Genes Encoding Neuronal Voltage-Gated Calcium Channel  
FILE REFERENCE: US App. 09/123,030  
CURRENT APPLICATION NUMBER: US/09/123,030  
CURRENT FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 40  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 1558  
TYPE: DNA  
ORGANISM: Murinae gen. sp.  
FEATURE:

NAME/KEY: 5'UTR  
LOCATION: (1)..(389)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (390)..(1361)  
FEATURE:  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (1361)..(1558)  
US-09-123-030-7

Query Match 76.0%; Score 15.2; DB 4; Length 1558;  
Best Local Similarity 85.0%; Pred. No. 1.5e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
DB 1350 TGGTCGGCGGTGGCTGTG 1331

## RESULT 15

US-08-242-689-1  
Sequence 1: Application US/08242689

Patent No. 5645998

## GENERAL INFORMATION:

APPLICANT: Atkinson, Mark A.  
APPLICANT: MacLaren, No. 56459981 K.  
APPLICANT: Kastern, William  
TITLE OF INVENTION: Methods and Compositions for Early  
Detection and Treatment of Insulin Dependent Diabetes  
TITLE OF INVENTION: Mellitus  
NUMBER OF SEQUENCES: 7

## CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/242,689  
FILING DATE:

## CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/746,443  
FILING DATE: 16-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/569,324  
FILING DATE: 17-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/427,051  
FILING DATE: 25-OCT-1989

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/283,633  
FILING DATE: 13-DEC-1988  
ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF/S&S-18.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

LENGTH: 2265 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1875  
US-08-242-689-1

Query Match 76.0%; Score 15.2; DB 1; Length 2265;  
Best Local Similarity 60.0%; Pred. No. 1.5e+02;  
Matches 12; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
DB 1008 UGGAGCUGCACUUGGCUUUG 1027

Search completed: July 8, 2003, 09:30:59  
Job time : 38.6598 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 115,512 Seconds  
(without alignments)  
273,390 Million cell updates/sec

Title: US-09-647-780A-5  
Perfect score: 20  
Sequence: 1 tggagcggcagttgctgtg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications, NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.4	82.0	22452	9	US-09-764-868-1487, Ap
2	16.4	82.0	22452	9	US-09-764-868-1489
3	15.8	79.0	3348	10	US-09-781-558-1
4	15.8	79.0	4197	10	US-09-957-683-1
5	15.8	79.0	5746	9	US-09-764-891-5830
6	15.8	79.0	7808	9	US-10-205-823-159
7	15.8	79.0	34658	9	US-09-900-449A-3
8	15.8	79.0	74586	10	US-09-781-558-3
9	15.4	77.0	384	10	US-09-878-574-1723
10	15.4	77.0	9025608	9	US-10-156-761-1
11	15.2	76.0	296	10	US-09-960-352-6786
12	15.2	76.0	307	10	US-09-867-701-3025
13	15.2	76.0	332	9	US-09-918-995-12815
14	15.2	76.0	478	10	US-09-864-761-5691
15	15.2	76.0	482	9	US-09-918-995-26751
16	15.2	76.0	506	10	US-09-244-694-109
17	15.2	76.0	510	10	US-09-974-300-6034
18	15.2	76.0	516	10	US-09-917-800A-1224
19	15.2	76.0	531	9	US-10-198-846-9095

20	15.2	76.0	588	10	US-09-864-761-9074
21	15.2	76.0	810	9	US-10-156-761-4837
22	15.2	76.0	884	9	US-10-222-577-11
23	15.2	76.0	884	9	US-10-222-578-11
24	15.2	76.0	884	10	US-09-790-045-11
25	15.2	76.0	918	9	US-10-156-761-6700
26	15.2	76.0	1520	9	US-09-924-340-107
27	15.2	76.0	1520	9	US-09-992-600A-107
28	15.2	76.0	1520	9	US-10-000-489-107
29	15.2	76.0	1520	9	US-10-000-986-107
30	15.2	76.0	1527	9	US-09-938-842A-1223
31	15.2	76.0	1528	9	US-09-746-783-183
32	15.2	76.0	1554	9	US-10-037-270-802
33	15.2	76.0	1570	9	US-09-802-640-23
34	15.2	76.0	1600	9	US-10-171-581-86
35	15.2	76.0	1738	10	US-09-880-107-2118
36	15.2	76.0	2469	10	US-09-853-386-90
37	15.2	76.0	2799	9	US-10-205-951-94
38	15.2	76.0	2893	9	US-10-017-273A-4
39	15.2	76.0	2893	10	US-09-905-846-1
40	15.2	76.0	2975	9	US-10-017-273A-5
41	15.2	76.0	2975	10	US-09-905-846-5
42	15.2	76.0	3331	12	US-10-044-090-272
43	15.2	76.0	3408	9	US-10-128-714-342
44	15.2	76.0	3409	9	US-10-128-714-5342
45	15.2	76.0	3421	9	US-10-205-951-56

## ALIGNMENTS

RESULT 1  
US-09-764-868-1487  
; Sequence 1487, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P2332  
; CURRENT APPLICATION NUMBER: US/09764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1487  
; LENGTH: 22452  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-764-868-1487

Query Match 82.0%; Score 16.4; DB 9; Length 22452;  
Best Local Similarity 94.4%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GAGCGGCGAGTGGCTGTG 20  
Db 22357 GAGCGGCGAGTGGCTGTG 22374  
|||||

RESULT 2  
US-09-764-868-1489  
; Sequence 1489, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: P2332  
; CURRENT APPLICATION NUMBER: US/09764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1489

LENGTH: 22452  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-868-1489

Query Match 82.0%; Score 16.4; DB 9; Length 22452;  
Best Local Similarity 94.4%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 GAGCGGCAGTTGGCTGTG 20  
|||||  
Db 22357 GAGCGGCAGTTGGCTGTG 22374

## RESULT 3

US-09-781-558-1  
Sequence 1, Application US/09781558  
Patent No. US20020034778A1  
GENERAL INFORMATION:  
APPLICANT: Beasley et al.  
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: AND USES THEREOF  
FILE REFERENCE: CL000756  
CURRENT APPLICATION NUMBER: US/09/781,558  
CURRENT FILING DATE: 2001-02-13  
PRIOR APPLICATION NUMBER: US 60/199,811  
PRIOR FILING DATE: 2000-04-26  
PRIOR APPLICATION NUMBER: US 09/641,426  
PRIOR FILING DATE: 2000-08-18  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 3348  
TYPE: DNA  
ORGANISM: HUMAN  
US-09-781-558-1

Query Match 79.0%; Score 15.8; DB 10; Length 3348;  
Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GGAGCGCAGTTGGCTGTG 20  
|||||  
Db 3273 GGAGCTGCAGCTGGCTGTG 3291

## RESULT 4

US-09-957-683-1  
Sequence 1, Application US/09957683  
Patent No. US20020123098A1  
GENERAL INFORMATION:  
APPLICANT: Curtis, Rory A. J.  
TITLE OF INVENTION: 55063, a No. US20020123098A1 Human NMDA Family Member  
TITLE OF INVENTION: and Uses Thereof  
FILE REFERENCE: MNI-191  
CURRENT APPLICATION NUMBER: US/09/957,683  
CURRENT FILING DATE: 2001-09-19  
PRIOR APPLICATION NUMBER: 60/233,537  
PRIOR FILING DATE: 2000-09-19  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 4197  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-957-683-1

Query Match 79.0%; Score 15.8; DB 10; Length 4197;  
Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GGAGCGCAGTTGGCTGTG 20

Db 3273 GGAGCTGCAGCTGGCTGTG 3291  
|||||

## RESULT 5

US-09-764-891-5830  
Sequence 5830, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 5830  
LENGTH: 5746  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-5830

Query Match 79.0%; Score 15.8; DB 9; Length 5746;  
Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGAGCGCAGTTGGCTGT 19  
|||||

Db 1835 TGGAGCGCAGTTGGCTGT 1853

## RESULT 6

US-10-205-823-159  
Sequence 159, Application US/10205823  
Publication No. US20030108963A1  
GENERAL INFORMATION:  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Endege, Wilson O.  
APPLICANT: Gannavarapu, Manjula  
APPLICANT: Gorbacheva, Bella  
APPLICANT: Hoersch, Sebastian  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Wansley, Angela M.  
APPLICANT: Glatz, Karen  
APPLICANT: Zhao, Xumei  
APPLICANT: Anderson, Dustin  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
TITLE OF INVENTION: THERAPY OF PROSTATE CANCER  
FILE REFERENCE: MRI-044  
CURRENT APPLICATION NUMBER: US/10/205,823  
CURRENT FILING DATE: 2002-07-25  
PRIOR APPLICATION NUMBER: 60/307,982  
PRIOR FILING DATE: 2001-07-25  
PRIOR APPLICATION NUMBER: 60/314,356  
PRIOR FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/325,020  
PRIOR FILING DATE: 2001-09-25  
PRIOR APPLICATION NUMBER: 60/341,746  
PRIOR FILING DATE: 2001-12-12  
PRIOR APPLICATION NUMBER: 60/362,158  
PRIOR FILING DATE: 2002-03-05  
NUMBER OF SEQ ID NOS: 455  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 159  
LENGTH: 7808  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-205-823-159

Query Match

79.0%; Score 15.8; DB 9; Length 7808;

Best Local Similarity 89.5%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTGTG 20  
||||| ||||| ||||| |||||  
Db 3893 GGAGCTGCAGCTGGCTGTG 3911

## RESULT 7

US-09-900-449A-3  
; Sequence 3, Application US/09900449A  
; Publication No. US20030040616A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHONG, Jenny et al.  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND  
; FILE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CLO01271  
; CURRENT APPLICATION NUMBER: US/09/900,449A  
; CURRENT FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 34668  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-900-449A-3

Query Match 79.0%; Score 15.8; DB 9; Length 34668;  
Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTGTG 20  
||||| ||||| ||||| |||||  
Db 17989 GGAGGGCAGTGGGTGTG 18007

## RESULT 8

US-09-781-558-3  
; Sequence 3, Application US/09781558  
; Patent No. US20020034778A1  
; GENERAL INFORMATION:  
; APPLICANT: Beasley et al.  
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
; FILE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: CLO00756  
; CURRENT APPLICATION NUMBER: US/09/781,558  
; CURRENT FILING DATE: 2001-02-13  
; PRIOR APPLICATION NUMBER: US 60/199,811  
; PRIOR FILING DATE: 2000-04-26  
; PRIOR APPLICATION NUMBER: US 09/641,426  
; PRIOR FILING DATE: 2000-08-18  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 74586  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(74586)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-781-558-3

Query Match 79.0%; Score 15.8; DB 10; Length 74586;  
Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTGTG 20  
||||| ||||| ||||| |||||  
Db 69665 GGAGCTGCAGCTGGCTGTG 69683

## RESULT 9

US-09-878-574-1723  
; Sequence 1723, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 1723  
; LENGTH: 384  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: LIB3028-034-Q1-B1-D2  
US-09-878-574-1723

Query Match 77.0%; Score 15.4; DB 10; Length 384;  
Best Local Similarity 94.1%; Pred. No. 3.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCT 17  
||||| ||||| ||||| |||||  
Db 170 TGGAGCGGCAGTTGTCT 186

## RESULT 10

US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMDRA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 77.0%; Score 15.4; DB 9; Length 9025608;  
Best Local Similarity 94.1%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTTGGCTG 18  
||||| ||||| ||||| |||||  
Db 3297015 GGAGCGGCAGTTGCTG 3297031

## RESULT 11

US-09-960-352-6786/c  
; Sequence 6786, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 6786

; LENGTH: 296

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 29-LIB34-015-Q1-E1-H1

US-09-960-352-6786

Query Match 76.0%; Score 15.2; DB 10; Length 296;

Best Local Similarity 85.0%; Pred. No. 3.9e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20

||||| | ||||| |

Db 217 TGGAGCGTCGTTGGCTTTG 198

||||| | ||||| |

## RESULT 12

US-09-867-701-3025/c

; Sequence 3025, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Aglate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3025

; LENGTH: 307

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-867-701-3025

Query Match 76.0%; Score 15.2; DB 10; Length 307;

Best Local Similarity 85.0%; Pred. No. 3.9e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20

||||| | ||||| |

Db 56 TGGGGGGGCGAGTGGCTGTG 37

||||| | ||||| |

## RESULT 13

US-09-918-995-12815

; Sequence 12815, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 12815

; LENGTH: 332

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-918-995-12815

Query Match 76.0%; Score 15.2; DB 9; Length 332;

Best Local Similarity 85.0%; Pred. No. 3.8e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20

||||| | ||||| |

Db 11 TGGAGCCCGAGTGTGCTGTG 30

||||| | ||||| |

## RESULT 14

US-09-864-761-5691/c

; Sequence 5691, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL

; FILE REFERENCE: Aesomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117

; SOFTWARE: Anncmax Sequence Listing Engine vers. 1.1

; SEQ ID NO 5691

; LENGTH: 478

; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC004752.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 21  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 18  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 13  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 15  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 15  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 17  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 16  
US-09-864-761-5691

Query Match 76.0%; Score 15.2; DB 10; Length 478;  
Best Local Similarity 85.0%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
||||| ||||||| |||||  
DB 314 TGGAGTGGCAGTTGCTGAG 295

RESULT 15  
US-09-918-995-26751  
; Sequence 26751, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26751  
; LENGTH: 482  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-26751

Query Match 76.0%; Score 15.2; DB 9; Length 482;  
Best Local Similarity 85.0%; Pred. No. 3.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
||||| ||||||| |||||||  
DB 159 TGGAGCGGATGCTGTG 178

Search completed: July 9, 2003, 02:21:46  
Job time : 130.562 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1013.95 seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-5

Perfect score: 20

Sequence: 1 tggagcggcagtgctgtg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:  
1: em\_estha.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pin.\*  
21: em\_gss\_vit.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	600	13	BI989738 4044-75 M
2	17.4	87.0	509	13	BJ173933
3	17.4	87.0	538	13	BJ184876
4	17.4	87.0	565	17	A2367683
5	17.4	87.0	601	13	BJ187554
6	17.4	87.0	660	14	BQ521069 NISC.n108

7	17.4	87.0	786	17	AQ292016
8	17.4	87.0	806	17	AQ688530
9	17	85.0	397	10	BE469456
10	16.8	84.0	214	9	AA093253
11	16.8	84.0	265	12	BF891199
12	16.8	84.0	316	17	AZ447470
13	16.8	84.0	332	17	AQ443490
14	16.8	84.0	396	12	BF511828
15	16.8	84.0	418	10	AW898919
16	16.8	84.0	455	12	BG406657
17	16.8	84.0	472	17	AZ647506
18	16.8	84.0	476	17	AQ407744
19	16.8	84.0	479	13	BI129332
20	16.8	84.0	482	10	AV742701
21	16.8	84.0	511	14	BM750988
22	16.8	84.0	528	17	AZ625044
23	16.8	84.0	534	10	BB704298
24	16.8	84.0	567	10	AW638441
25	16.8	84.0	617	13	BI354908
26	16.8	84.0	620	13	BI939138
27	16.8	84.0	634	9	AA950083
28	16.8	84.0	661	9	AA948928
29	16.8	84.0	661	17	BH521881
30	16.8	84.0	686	17	CNS04LJF
31	16.8	84.0	707	13	BI117591
32	16.8	84.0	716	11	BC032146
33	16.8	84.0	724	12	BF504506
34	16.8	84.0	737	12	BF494912
35	16.8	84.0	781	12	BF486440
36	16.8	84.0	801	12	BG398947
37	16.8	84.0	896	17	CNS04G66
38	16.8	84.0	928	9	AA857841
39	16.8	84.0	1039	17	CNS04IMF
40	16.8	84.0	1062	17	CNS02URU
41	16.4	82.0	170	12	BF887478
42	16.4	82.0	232	14	BQ344355
43	16.4	82.0	233	12	BF887611
44	16.4	82.0	250	14	BQ345871
45	16.4	82.0	258	12	BF765213

# ALIGNMENTS

RESULT 1  
BI989738 4044-75 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
LOCUS mRNA sequence.  
DEFINITION  
BI989738  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

600 bp  
4044-75 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
mRNA sequence.  
GI:17960740  
Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 600)  
Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
21671825  
Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers  
1..600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"



ACCESSION A2367683  
 VERSION A2367683.1 GI:10481479  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 565)  
 AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112 USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0117 row: M column: 01  
 Seq primer: CACAGGACAGCAGTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 565.

FEATURES  
 source  
 1. 565  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UGC1M0117M01"  
 /clone.lib="Mouse 10kb plasmid UGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: PWD42hv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 168 a 127 c 136 g 134 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 565;  
 Best Local Similarity 94.7%; Pred. No. 1.7e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGAGCGGACGTGGCTGTG 20  
 ||||| ||||| ||||| ||||| |||||  
 Db 270 GGAAGCGGACGTGGCTGTG 288

RESULT 5  
 BJI87554 601 bp mRNA linear EST 24-JAN-2002  
 LOCUS BJI87554 normalized full length cDNA library, chloronemata, caulonemata and malformed buds Physcomitrella patens subsp. patens  
 DEFINITION

CDNA clone pphb41106 5', mRNA sequence.  
 BJI87554  
 BJI87554.1 GI:18355495  
 EST.  
 SOURCE Physcomitrella patens subsp. patens.  
 ORGANISM Physcomitrella patens subsp. patens.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta; Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.

REFERENCE 1 (bases 1 to 601)  
 AUTHORS Fujita, T., Shin-i, T., Seki, M., Kamiya, A., Uchiyama, I., Nishiyama, T., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M.

TITLE Comparison of the moss Physcomitrella patens genome with flowering plants genome

JOURNAL Unpublished (2002)  
 COMMENT Contact: Tadasu Shin-i  
 Center For Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshini@genes.nig.ac.jp

A backbone of the vector is pBluescript II, that was in vivo excised from a modified LPS phage vector (Mo bi Tec, Germany). XhoI digested-5' end of cDNA is ligated to SalI site of the vector, and the BamHI digested-3' end including poly-A tail is ligated to BamHI site of the vector. cDNA insert could be amplified with conventional T7 and T3 primers. This normalized full-length cDNA library was generated basically according to the method described in Genome Research 10, 1617-1630 (2000). Carninci, P. et al. Protonemata were blended by the POLYTRON, and then cultivated on the BCD medium containing 0.5uM BA (benzylaminopurine) for 8 to 13 days under the continuous light.

FEATURES  
 source  
 1. 601  
 /organism="Physcomitrella patens subsp. patens"  
 /db\_xref="taxon:145481"  
 /clone="pphb41106"  
 /clone.lib="normalized full length cDNA library, chloronemata, caulonemata and malformed buds"  
 /tissue\_type="mixture of chloronemata, caulonemata and malformed buds"

BASE COUNT 84 a 174 c 145 g 198 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 13; Length 601;  
 Best Local Similarity 94.7%; Pred. No. 1.8e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGAGCGGACGTGGCTGTG 20  
 ||||| ||||| ||||| ||||| |||||  
 Db 454 GGAGCGGACGTGGCTGTG 472

RESULT 6  
 BQ521069 660 bp mRNA linear EST 10-JUN-2002  
 LOCUS BQ521069  
 DEFINITION NISC\_n108h11.v1 NICHG\_XGC\_Emb7 Silurana tropicalis cDNA clone  
 IMAGE:5335941.5', mRNA sequence.

ACCESSION BQ521069  
 VERSION BQ521069.1 GI:21379938  
 KEYWORDS EST.  
 SOURCE western clawed frog.  
 ORGANISM Silurana tropicalis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.

REFERENCE 1 (bases 1 to 660)  
 AUTHORS NIH-XCG http://image.llnl.gov/image/html/xenopuslib\_info.shtml  
 TITLE National Institute of Child Health and Human Development, National Cancer Institute, Xenopus Gene Collection  
 JOURNAL Unpublished (2002)  
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: c9apbs-r@mail.nih.gov  
 CDNA Library Preparation:  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 info@image.llnl.gov  
 Plate: LLAM11852 row: P column: 22  
 Seq primer: M13Rpi reverse primer (ABI).  
 Location/Qualifiers

#### FEATURES

source

1. .660  
 /organism="Silurana tropicalis"  
 /db\_xref="taxon:8364"  
 /clone="IMAGE:5335941"  
 /clone\_lib="NICHD\_XGC\_Emb7"  
 /tissue\_type="tailbud"  
 /dev\_stage="embryo, stages 20-27"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Vector: pCMV-SPORT6.ccdB; Site\_1: NotI; Site\_2:  
 EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average  
 insert size 2.1 kb. Constructed by Invitrogen. Note: This  
 is a Xenopus Gene Collection (XGC) library."  
 141 a 170 c 198 g 151 t

#### BASE COUNT

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 660;  
 Best Local Similarity 94.7%; Pred. No. 1.8e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGAGCGGCAGTGGCTGTG 20

Db 177 GGGCGGCAGTGGCTGTG 195

#### RESULT 7

AQ292016

#### LOCUS

DEFINITION nbx0041E05f CUGI Rice BAC Library Oryza sativa genomic clone  
 nbx0041E05f, DNA sequence.

#### ACCESSION

AQ292016

#### VERSION

AQ292016.1 GI:3953210

#### KEYWORDS

GSS.

#### SOURCE

Oryza sativa.

#### ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

#### REFERENCE

1 (bases 1 to 786)

#### AUTHORS

Wing, R.A. and Dean, R.A.

#### TITLE

A BAC End Sequencing Framework to Sequence the Rice Genome

#### JOURNAL

Unpublished (1998)

#### COMMENT

Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: TAATAGACTACTATAGG  
 Class: BAC ends  
 High quality sequence stop: 416.  
 Location/Qualifiers

#### FEATURES

source

1. .786  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbx0041E05f"  
 /clone\_lib="CUGI Rice BAC Library"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:

HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

#### BASE COUNT

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 786;  
 Best Local Similarity 94.7%; Pred. No. 1.9e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTGGCTGT 19

Db 465 TGGAGCGGCAGTGGCTGT 483

#### RESULT 8

AQ688530

#### LOCUS

DEFINITION nbx0077H20r CUGI Rice BAC Library Oryza sativa genomic clone  
 nbx0077H20r, DNA sequence.

#### ACCESSION

AQ688530

#### VERSION

AQ688530.1 GI:5329698

#### KEYWORDS

GSS.

#### SOURCE

Oryza sativa.

#### ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

#### REFERENCE

1 (bases 1 to 806)

#### AUTHORS

Wing, R.A. and Dean, R.A.

#### TITLE

A BAC End Sequencing Framework to Sequence the Rice Genome

#### JOURNAL

Unpublished (1998)

#### COMMENT

Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: GGAACAGCTATGACCATG  
 Class: BAC ends  
 High quality sequence stop: 196.  
 Location/Qualifiers

#### FEATURES

source

1. .806  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbx0077H20r"  
 /clone\_lib="CUGI Rice BAC Library"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:

HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant

(2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 165 a 198 c 203 g 230 t 10 others  
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 806;  
Best Local Similarity 94.7%; Pred. No. 1.9e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGT 19  
|||||  
Db 467 TGGAGCGGCAGTTGGCTGT 485

RESULT 9  
LOCUS BE469456/c  
DEFINITION IPHdk01836 Head kidney cDNA library, Ictalurus punctatus cDNA 5', similar to Ribosomal protein L36a, mRNA sequence.  
ACCESSION BE469456  
VERSION BE469456.1 GI:9559947  
KEYWORDS EST.  
SOURCE channel catfish.  
ORGANISM Ictalurus punctatus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.  
REFERENCE Cao, D., Kocabas, A., Ju, Z., Karsi, A., Li, P., Patterson, A. and Liu, Z.J.  
AUTHORS

TITLE Transcriptome analysis of channel catfish (Ictalurus punctatus): Gene cataloging and profiling from the head kidney  
JOURNAL Unpublished (2000)  
COMMENT Contact: Liu ZJ  
The Fish Molecular Genetics and Biotechnology Laboratory, Department of Fisheries and Allied Aquacultures and Program of Cell and Molecular Biosciences  
Auburn University  
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA  
Tel: 334 844 4054  
Fax: 334 844 9208  
Email: zliu@acesag.auburn.edu  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .397  
/organism="Ictalurus punctatus"  
/db\_xref="taxon:7998"  
/clone\_lib="Head kidney cDNA library"  
/note="Organ: Head kidney; Vector: pSport1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 126 a 87 c 113 g 71 t  
ORIGIN

Query Match 85.0%; Score 17; DB 10; Length 397;  
Best Local Similarity 100.0%; Pred. No. 2.3e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCT 17  
|||||  
Db 251 TGGAGCGGCAGTTGGCT 235

RESULT 10  
LOCUS AA093253  
DEFINITION 110282.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA 5', mRNA sequence.  
ACCESSION AA093253  
VERSION AA093253.1 GI:1638722  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 214)  
AUTHORS Liaw, C.C.  
TITLES CDNAs from fetal heart (1996)  
JOURNAL Unpublished (1996)  
COMMENT Contact: Liew CC  
Brigham and Women's Hospital  
Harvard Medical School  
75 Francis St. Boston, MA 02115, USA  
Tel: 6177328915  
Fax: 617750995  
Email: cliaw@rics.bwh.harvard.edu  
PCR Primers  
FORWARD: 5' GCCAAGCTCGAATTACCTCACTAAAGG 3'  
BACKWARD: 5' CCAGTGAATTGATACGACTCACTATAGGCG 3'  
Seq primer: 5' GAAATTAACCTCACTAAAGG 3'

FEATURES  
1..214  
Location/Qualifiers  
1. .214  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human fetal heart, Lambda ZAP Express"  
/lab\_host="E. coli XL1-Blue"  
/note="Vector: Lambda ZAP Express; Site\_1: EcoRI; Site\_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). CDNA was synthesized using a XhoI-oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda ZAP Express."

BASE COUNT 33 a 82 c 45 g 54 t  
ORIGIN

Query Match 84.0%; Score 16.8; DB 9; Length 214;  
Best Local Similarity 90.0%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGAGCGGCAGTTGGCTGTG 20  
|||||  
Db 27 TGGAGCGGCAGTTGGCTGTG 46

RESULT 11  
LOCUS BF891199  
DEFINITION PM3-MT0110-191000-007-b11 MT0110 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF891199  
VERSION BF891199.1 GI:12282658  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 265)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663

## COMMENT

Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&t2=PM3-WT0110-  
 191000-007-bllst3-2000-10-19&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 69.

## FEATURES

## source

1. .265  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="WT0110"  
 /dev\_stage="Adult"  
 /note="Organ: marrow; Vector: puc18; Site: 1; Smail; Site 2:  
 Smail; A mini-library was made by cloning products derived  
 from ORSTES PCR (U.S. Letters Patent application No. 196  
 ,716 - Ludwig Institute for Cancer Research) profiles  
 into the puc 18 vector. Reverse transcription of tissue  
 mRNA and cDNA amplification were performed under low  
 stringency conditions."  
 BASE COUNT 63 a 59 c 63 g 80 t

## Query Match

Best Local Similarity 84.0%; Score 16.8; DB 12; Length 265;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20

Db 42 TGAAGTGGCGAGTTGGCTGTG 61

## RESULT 12

## A2447470

## LOCUS

AM02447470 Mouse 10kb plasmid UUGCLM library Mus musculus genomic  
 clone UUGCLM0244G11 R, DNA sequence.

## ACCESSION

A2447470

## VERSION

A2447470.1 GI:10599295

## KEYWORDS

GSS

## SOURCE

house mouse.

## ORGANISM

Mus musculus

## REFERENCE

1 (bases 1 to 316)

## AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly  
 M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
 and Wright,D., Weiss,R.

## TITLE

Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112 USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: dunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0244 row: G column: 11

Seq primer: CACACGAGAACAGCTATGACC

Class: plasmid end

High quality sequence stop: 316.

Location/Qualifiers

## source

1. .316  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCLM0244G11"  
 /clone\_lib="Mouse 10kb plasmid UUGCLM library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (GI47321141gb/AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adapted mouse DNA was annealed to  
 adapted vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."  
 BASE COUNT 74 a 48 c 72 g 122 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 17; Length 316;

Best Local Similarity 90.0%; Pred. No. 2.7e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20

Db 266 TGGAGCTTCAGTTGGCTGTG 285

## RESULT 13

## A04343490

## LOCUS

QST0078 Trypanosoma cruzi random genomic library Trypanosoma  
 cruzi genomic clone G10E17, DNA sequence.

## DEFINITION

A04343490

## ACCESSION

A04343490.3 GI:10140478

## VERSION

GSS

## KEYWORDS

Trypanosoma cruzi.

## SOURCE

Trypanosoma cruzi

## ORGANISM

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma; Schizotrypanum.

## REFERENCE

1 (bases 1 to 332)

## AUTHORS

Aguero,F., Verdun,R., Frasch,A.C.C. and Sanchez,D.O.

## TITLE

A random sequencing approach for the analysis of the trypanosoma  
 cruzi genome: general structure, large gene and repetitive DNA  
 families, and gene discovery

## JOURNAL

Genome Res. 10 (12), 1996-2005 (2000)

## MEDLINE

20588489

## COMMENT

On Sep 14, 2000 this sequence version replaced gi:9377133.

Contact: Sanchez D.O.

Instituto de Investigaciones Biologicas (Univ. Nac. de Gral

San Martin)

Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24

CP(1650) San Martin, Prov. de BS AS, Argentina

Tel: 54-11-4580-7255 ext 309

Fax: 54-11-4752-9639

Email: dsanchez@iib.unsam.edu.ar

Sequences were basecalled with phred and vector was masked with  
 crossmatch (see http://genome.washington.edu). Sequences were then  
 trimmed from both ends to remove low quality bases and masked  
 vector.

Seq primer: T7

Class: shotgun.

FEATURES  
source

Location/Qualifiers  
1. .332  
/organism="Trypanosoma cruzi"  
/strain="CL-Brener"  
/db\_xref="taxon:5693"  
/clone="G10817"  
/cell\_type="epimastigote"  
/note="Vector: pBS(-) (Stratagene); T. cruzi DNA was randomly sheared using a nebulizer and the 1 to 2 Kb range was gel purified and cloned into the dephosphorylated HincII site of the vector"

BASE COUNT 59 a 86 c 115 g 72 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 17; Length 332;  
Best Local Similarity 90.0%; Pred. No. 2.7e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
||||| ||||||| |||||  
DB 283 TGGAGCAGCGAGTTGGTTGTG 302

## RESULT 14

## BF511828

## LOCUS

UI-H-B14-abs-b-06-0-01.s1 NCI\_CGAP\_Sub8 Homo sapiens cDNA clone  
IMAGE:3088450 3', mRNA sequence.

## ACCESSION

## BF511828

## VERSION

## BF511828.1

## KEYWORDS

## EST.

## SOURCE

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.

## 1 (bases 1 to 396)

## NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

## National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

## Tumor Gene Index

## Unpublished (1997)

## Contact: Robert Strausberg, Ph.D.

## Email: cgabs@mail.nih.gov

## The sequence contained an oligo-dT track that was present in the

## oligonucleotide that was used to prime the synthesis of first

## strand cDNA and therefore this may represent a bonafide poly A

## tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:

## NCI-CGAP clone distribution information can be found through the

## I.M.A.G.E. Consortium/LLNL at:

## www.bio.llnl.gov/bbrp/image/image.html

## Seq primer: M13 Forward

## POLVA=Yes.

FEATURES  
source

Location/Qualifiers  
1. .396  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3088450"  
/clone\_lib="NCI\_CGAP\_Sub8"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; NCI\_CGAP\_Sub8 is a subtracted library derived from NCI\_CGAP\_Sub5. The NCI\_CGAP\_Sub8 library had 2.5 million recombinants. A single-stranded DNA preparation of NCI\_CGAP\_Sub5 was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI\_CGAP\_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI\_CGAP\_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population), NCI\_CGAP\_Sub6 (pool AIF-AJU, IMAGE ids 2728969-2733190; 25% of the driver population), and NCI\_CGAP\_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was

performed as previously described [Bonaldi, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG\_SEQ=None found"

BASE COUNT 88 a 66 c 79 g 163 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 396;  
Best Local Similarity 90.0%; Pred. No. 2.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGAGCGGCGAGTTGGCTGTG 20  
||||| ||||||| |||||||  
DB 232 TGAAGTCGCAGTTGGCTGTG 251

## RESULT 15

## AW898919

## LOCUS

CMO-NN0079-140400-334-all NN0079 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW898919  
VERSION AW898919.1  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

## Homo sapiens

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.

## 1 (bases 1 to 418)

## Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

## Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,

## Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,

## Brunstein, A., Geolivelira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

## , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

## Simpson, A.J.

## Shotgun sequencing of the human transcriptome with ORF expressed

## sequence tags

## Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## 20202663

## Contact: Simpson A.J.G.

## Laboratory of Cancer Genetics

## Ludwig Institute for Cancer Research

## Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

## Brazil

## Tel: +55-11-2704922

## Fax: +55-11-2707001

## Email: asimpson@ludwig.org.br

## This sequence was derived from the FAPESP/LICR Human Cancer Genome

## Project. This entry can be seen in the following URL

## (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-cmo-NN0079-140

## 400-334-all&amp;st3=2000-04-14&amp;st4=1)

## Seq primer: puc 18 forward

## High quality sequence stop: 418.

## Location/Qualifiers

## 1. .418

## /organism="Homo sapiens"

## /db\_xref="taxon:9606"

## /clone\_lib="NN0079"

## /dev stage="Adult"

## /note="Organ: nervous normal; Vector: puc18; Site\_1: SmaI;

## Site\_2: SmaI; A mini-library was made by cloning products

## derived from ORESTES PCR (U.S. Letters Patent application

## No. 196,716 - Ludwig Institute for Cancer Research)

## profiles into the puc 18 vector. Reverse transcription of

## tissue mRNA and cDNA amplification were performed under

## low stringency conditions."

## BASE COUNT 85 a 88 c 89 g 156 t

## ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 418;  
Best Local Similarity 90.0%; Pred. No. 2.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY . 1 TGGAGCGGCAGTTGGCTGTG 20  
||| ||| ||| ||| ||| |||  
Db 180 TGAAGTGGCAGTTGGCTGTG 199

Search completed: July 8, 2003, 09:21:12  
Job time : 1022.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 231.102 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-6  
Perfect score: 21  
Sequence: 1 agttccacttgaggcccatg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.in.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	21	100.0	21	6	AX014706
2	21	100.0	2765	6	AX014701
3	19.4	92.4	2583	10	AF302075
4	19.4	92.4	2601	10	AF157106
5	19.4	92.4	2652	10	AF302076
6	19.4	92.4	2694	10	AF302077
7	19.4	92.4	2892	10	AF157105
8	19.4	92.4	2925	6	AX033272
9	19.4	92.4	2925	10	AF176569
10	18.4	87.6	10208	1	AE000795
11	17.8	84.8	492	4	D42021S07
12	17.8	84.8	2300	8	AF439723
13	17.8	84.8	34864	9	U73649
14	17.8	84.8	84181	2	CNS08C8T
15	17.8	84.8	99239	2	CNS08C8V
16	17.8	84.8	100773	8	AF466201
17	17.8	84.8	137573	10	AL669884
18	17.8	84.8	137721	2	AC113928
19	17.8	84.8	141085	2	AC097250
20	17.8	84.8	149904	9	CNS01DSV
21	17.8	84.8	160222	9	AP002962
22	17.8	84.8	162021	2	AC119375
23	17.8	84.8	162691	9	CNS01DXA
24	17.8	84.8	184926	9	AP000757
25	17.8	84.8	198646	2	AC096281
26	17.8	84.8	200023	2	AC129297
27	17.8	84.8	202655	10	AC026682
28	17.8	84.8	208764	2	AL845263
29	17.8	84.8	239319	2	AC099583
30	17.4	82.9	63618	2	AC129959
31	17.4	82.9	85654	9	HSP373C6
32	17.4	82.9	134189	2	AC079550
33	17.4	82.9	157017	9	AL162389
34	17.4	82.9	157639	9	AC012031
35	17.4	82.9	160096	2	AC022657
36	17.4	82.9	177695	2	AC096154
37	17.4	82.9	192817	2	AC127063
38	17.4	82.9	193892	2	AC073037
39	17.4	82.9	200551	2	AC113734
40	17.4	82.9	226998	2	AC111095
41	17.4	82.9	237563	2	AC102954
42	17.4	82.9	247196	2	AC073822
43	17.4	82.9	250615	2	AC073814
44	17	81.0	707	9	HS437796
45	17	81.0	174953	2	AC094732

ALIGNMENTS

RESULT 1	AX014706	Sequence 6 from Patent WO9553077.	21 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX014706					
DEFINITION	AX014706					
ACCESSION	AX014706					
VERSION	AX014706.1	GI:10040979				
KEYWORDS		synthetic construct.				
SOURCE		synthetic construct.				
ORGANISM		artificial sequences.				
REFERENCE		1 (bases 1 to 21)				
AUTHORS		Schwartz,J.C., Gros,C., Ouilmet,T., Rose,C., Bonhomme,M.C. and Facchinetti,P.				
TITLE		Novel nep ii membrane metalloprotease and its use for screening inhibitors useful in therapy				

JOURNAL Patent: WO 9953077-A 6 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source 1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
BASE COUNT 3 a 7 c 6 g 5 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21  
Db 1 AGTCCCACTGGGGCCCATG 21  
RESULT 2  
AX014701/c  
LOCUS AX014701 2765 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from Patent WO9953077.  
ACCESSION AX014701  
VERSION AX014701.1 GI:10040975  
KEYWORDS  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 2765)  
Schwartz,J.C., Gros,C., Oumet,T., Rose,C., Bonhomme,M.C. and  
Facchinetti,P.  
Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source 1..2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107..2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSESVGMERADNCGRRRLGVEGGLLVLLTLLMGAIVTL  
GVFYSIQKQLPILLSLHSHVSHRTYVVRVLRDSSQSDICTTSCVIAAARILQNM  
QSKPKDFYACGGWLRHVIPETNSRYSVDFILRDELEVILKGVLEDSVGRHPA  
VEKATLYRSCMGVSVIEKRDSEPLLVNIGGWPVAMDKWNETMGPKWLEQLAV  
LNSQFNRRVLDLFIWDDQNSRHVYVIDOPTLGMPSREYFEKDSHVRVAVLOEM  
TSVATMLRDLPLGERTLQVEMAOVLHETHLANATVPQEKRDVYALYRGLLEE  
LQERFLGFWNTLFIQNLVSQVELLPNEEVYVIGVPIYLENLEETIDVPAQTLQN  
LQVRLVLDIGLSQFKEARVDYRKALYGTMTVEVRKCVSVYNSNMESAVGSLY  
IKRAFSDKDSISVELIEKIRSVFVNDLNMWDESKKAQKALNIRQIGYPDY  
ILEDNRHLDSEYSLFSEDLYFENGQNLKNAQSLKLRKVDQNLWIGAAVY  
NIFYSPNRLVFPAGILOPFFSKDOPOALNFGGIGMWIGVHEITHGFDNDNGFQKN  
GNMIDWNSFSARHFOQSCMIYQYSNFWELADNONVNGFSLGENIADNGVQRA  
YKALQHLAEGRQORLPGLNLTAAQLFFINVAQVWCGSYRPEFAIQSIKTDVHSPK  
IRVLGSLQNLPGFSEAFHCPGSPMPHMRKRIW"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21

JOURNAL Patent: WO 9953077-A 6 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source 1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
BASE COUNT 3 a 7 c 6 g 5 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21  
Db 1 AGTCCCACTGGGGCCCATG 21  
RESULT 2  
AX014701/c  
LOCUS AX014701 2765 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from Patent WO9953077.  
ACCESSION AX014701  
VERSION AX014701.1 GI:10040975  
KEYWORDS  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 2765)  
Schwartz,J.C., Gros,C., Oumet,T., Rose,C., Bonhomme,M.C. and  
Facchinetti,P.  
Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source 1..2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107..2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSESVGMERADNCGRRRLGVEGGLLVLLTLLMGAIVTL  
GVFYSIQKQLPILLSLHSHVSHRTYVVRVLRDSSQSDICTTSCVIAAARILQNM  
QSKPKDFYACGGWLRHVIPETNSRYSVDFILRDELEVILKGVLEDSVGRHPA  
VEKATLYRSCMGVSVIEKRDSEPLLVNIGGWPVAMDKWNETMGPKWLEQLAV  
LNSQFNRRVLDLFIWDDQNSRHVYVIDOPTLGMPSREYFEKDSHVRVAVLOEM  
TSVATMLRDLPLGERTLQVEMAOVLHETHLANATVPQEKRDVYALYRGLLEE  
LQERFLGFWNTLFIQNLVSQVELLPNEEVYVIGVPIYLENLEETIDVPAQTLQN  
LQVRLVLDIGLSQFKEARVDYRKALYGTMTVEVRKCVSVYNSNMESAVGSLY  
IKRAFSDKDSISVELIEKIRSVFVNDLNMWDESKKAQKALNIRQIGYPDY  
ILEDNRHLDSEYSLFSEDLYFENGQNLKNAQSLKLRKVDQNLWIGAAVY  
NIFYSPNRLVFPAGILOPFFSKDOPOALNFGGIGMWIGVHEITHGFDNDNGFQKN  
GNMIDWNSFSARHFOQSCMIYQYSNFWELADNONVNGFSLGENIADNGVQRA  
YKALQHLAEGRQORLPGLNLTAAQLFFINVAQVWCGSYRPEFAIQSIKTDVHSPK  
IRVLGSLQNLPGFSEAFHCPGSPMPHMRKRIW"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21

JOURNAL Patent: WO 9953077-A 6 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source 1..21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
BASE COUNT 3 a 7 c 6 g 5 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21  
Db 1 AGTCCCACTGGGGCCCATG 21  
RESULT 2  
AX014701/c  
LOCUS AX014701 2765 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from Patent WO9953077.  
ACCESSION AX014701  
VERSION AX014701.1 GI:10040975  
KEYWORDS  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 2765)  
Schwartz,J.C., Gros,C., Oumet,T., Rose,C., Bonhomme,M.C. and  
Facchinetti,P.  
Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source 1..2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107..2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSESVGMERADNCGRRRLGVEGGLLVLLTLLMGAIVTL  
GVFYSIQKQLPILLSLHSHVSHRTYVVRVLRDSSQSDICTTSCVIAAARILQNM  
QSKPKDFYACGGWLRHVIPETNSRYSVDFILRDELEVILKGVLEDSVGRHPA  
VEKATLYRSCMGVSVIEKRDSEPLLVNIGGWPVAMDKWNETMGPKWLEQLAV  
LNSQFNRRVLDLFIWDDQNSRHVYVIDOPTLGMPSREYFEKDSHVRVAVLOEM  
TSVATMLRDLPLGERTLQVEMAOVLHETHLANATVPQEKRDVYALYRGLLEE  
LQERFLGFWNTLFIQNLVSQVELLPNEEVYVIGVPIYLENLEETIDVPAQTLQN  
LQVRLVLDIGLSQFKEARVDYRKALYGTMTVEVRKCVSVYNSNMESAVGSLY  
IKRAFSDKDSISVELIEKIRSVFVNDLNMWDESKKAQKALNIRQIGYPDY  
ILEDNRHLDSEYSLFSEDLYFENGQNLKNAQSLKLRKVDQNLWIGAAVY  
NIFYSPNRLVFPAGILOPFFSKDOPOALNFGGIGMWIGVHEITHGFDNDNGFQKN  
GNMIDWNSFSARHFOQSCMIYQYSNFWELADNONVNGFSLGENIADNGVQRA  
YKALQHLAEGRQORLPGLNLTAAQLFFINVAQVWCGSYRPEFAIQSIKTDVHSPK  
IRVLGSLQNLPGFSEAFHCPGSPMPHMRKRIW"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21

Db 741 AGTCCCACTGGGGCCCATG 721  
RESULT 3  
AF302075/c  
LOCUS AF302075 2583 bp mRNA linear ROD 11-JUN-2001  
DEFINITION Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.  
ACCESSION AF302075  
VERSION AF302075.1 GI:10505359  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2583)  
Shirotani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,  
Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,  
Iwatsubo,T. and Saido,T.C.  
Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most  
rapidly and efficiently among thiorphan- and  
phosphoramidon-sensitive endopeptidases  
J. Biol. Chem. 276 (24), 21895-21901 (2001)  
11278416  
2 (bases 1 to 2583)  
Shirotani,K. and Saido,T.C.  
Direct submission  
Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain  
Science Institute, RIKEN, 2-1 Hirotsawa, Wako-shi, Saitama  
351-0198, Japan  
FEATURES Location/Qualifiers  
source 1..2583  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
25..2253  
/note="endopeptidase"  
/codon\_start=1  
/product="neprilysin-like peptidase alpha"  
/protein\_id="AAG18446.1"  
/db\_xref="GI:10505360"  
/translation="MVERAGWCRKKSPGVEYGLMVLVLLLLGAVITLVGVYSIALRD  
SSLSKSDICTTSCVIAAARILEMDDSRNCFNYACGGWLRHVIPETNSRYSVF  
DILRDELEVILKGVLEDSQHRPAVEAKTLYRSCMGVSVIEKRDSEPLLVNIGGWPVAMDKWNETMGPKWLEQLAV  
LNSQFNRRVLDLFIWDDQNSRHVYVIDOPTLGMPSREYFEKDSHVRVAVLOEM  
TSVATMLRDLPLGERTLQVEMAOVLHETHLANATVPQEKRDVYALYRGLLEE  
LQERFLGFWNTLFIQNLVSQVELLPNEEVYVIGVPIYLENLEETIDVPAQTLQN  
LQVRLVLDIGLSQFKEARVDYRKALYGTMTVEVRKCVSVYNSNMESAVGSLY  
IKRAFSDKDSISVELIEKIRSVFVNDLNMWDESKKAQKALNIRQIGYPDY  
ILEDNRHLDSEYSLFSEDLYFENGQNLKNAQSLKLRKVDQNLWIGAAVY  
NIFYSPNRLVFPAGILOPFFSKDOPOALNFGGIGMWIGVHEITHGFDNDNGFQKN  
GNMIDWNSFSARHFOQSCMIYQYSNFWELADNONVNGFSLGENIADNGVQRA  
YKALQHLAEGRQORLPGLNLTAAQLFFINVAQVWCGSYRPEFAIQSIKTDVHSPK  
IRVLGSLQNLPGFSEAFHCPGSPMPHMRKRIW"  
BASE COUNT 665 a 667 c 736 g 515 t  
ORIGIN  
Query Match 92.4%; Score 19.4; DB 10; Length 2583;  
Best Local Similarity 95.2%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 AGTCCCACTGGGGCCCATG 21  
Db 560 AGTCCCACTGGGGCCCATG 540  
RESULT 4  
AF157106/c  
LOCUS AF157106 2601 bp mRNA linear ROD 25-NOV-1999  
DEFINITION Mus musculus soluble secreted endopeptidase delta mRNA,  
alternatively spliced product, complete cds.  
ACCESSION AF157106

AFI57106.1 GI:6467400

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2601)  
Ikeda,K., Emoto,N., Raharjo,S.B., Nurfantari,Y., Saiki,K.,  
Yokoyama,M. and Matsuo,M.  
Molecular identification and characterization of novel  
membrane-bound metalloprotease, the soluble secreted form of which  
hydrolyzes a variety of vasoactive peptides  
J. Biol. Chem. 274 (45), 32469-32477 (1999)

JOURNAL MEDLINE PUBLISHED  
REFERENCE AUTHORS TITLE JOURNAL  
Submitted (08-JUN-1999) International Center for Medical Research,  
Kobe University School of Medicine, 7-5-1 Kusunoki, Chuoh, Kobe  
6500017, Japan

FEATURES Location/Qualifiers  
source 1...2601  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
59..2287  
CDS .  
/note="SEP(delta); metalloprotease; alternatively spliced"  
/codon\_start=1  
/product="soluble secreted endopeptidase delta"  
/protein\_id="AAFI3153.1"  
/db\_xref="GI:6467401"

/translation="MVRRAGWCRKKSPGVEYLMLVLLLLGAIVTLGVFYSIGLKD  
SLSDICTPSCVTIAARILENDQSRNCFNYACGGWLHRHVIPETNSRYSVF  
DLRLEVLKGLEDTSOHPRAVERAKTLYRSQMNOVSVEKRDEPELLSKLVKMG  
CWPAVDKNFTMGKLWEKEROLAVINOFNRVLDLFWDNDQSSRHVIIDQPT  
LMPSRETYFDQDNHNVKVAPEMTSATMLRKDONLSKESANVRMAEMAEVLELT  
HLANATVPQEKRHDTHLANATVPQEKRHDTHLANATVPQEKRHDTHLANATVPQ  
VVGYPILENEDIDSARTMONYLWRVLDRIGLSQRFEARDVDYRKALYG  
TVEEYRWECVSYNMESAVGLYIKRAFSDKSSTRELIERISRVFNLDDELN  
WNDESCKKAOKAMNIROIQPYDILEDNNKHDEYSVLTFFEDLYFENGLONLK  
NMASRLKLRKVQNLIICAAVNAPYSPNRNQIVFPAGILOPPPKDPOSLN  
FGCGIWGHETHGDNGRNFRDKGNMDWSNFSAHFQQSQCMCIYQVNSWE  
LADNONGFSGLSIENADNGVRQAIRKAYLRWLADGGKDRPLGLNTLYAQILFYNI  
AQCWGCSYRPFPAQSITDVHSFKYRVLGSQNLPGFSEAFHGPRGSPMPMKRCR  
IW"

BASE COUNT 655 a 681 c 748 g 517 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 2601;  
Best Local Similarity 95.2%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTTCCACTTGCGGCCCATG 21  
||||||| |||||  
Db 594 AGTTCCACTTGCGGCCCATG 574

RESULT 5  
AF302076/c  
LOCUS AF302076 2652 bp mRNA linear ROD 11-JUN-2001  
DEFINITION Mus musculus neprilysin-like peptidase beta mRNA, complete cds.  
ACCESSION AF302076  
VERSION AF302076.1 GI:10505361  
KEYWORDS SOURCE  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2652)  
Shirotsani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,  
Maruyama,K., Kiryu-Seo,S., Miyama,H., Iwata,H., Tomita,T.,  
Iwasubo,T. and Saido,T.C.  
Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most  
rapidly and efficiently among thiorphan- and phosphoramidon-sensitive endopeptidases

JOURNAL MEDLINE PUBLISHED  
REFERENCE AUTHORS TITLE JOURNAL  
Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain  
Science Institute, RIKEN, 2-1 Hirotsawa, Wako-shi, Saitama  
351-0198, Japan

FEATURES Location/Qualifiers  
source 1...2652  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
25..2322  
CDS .  
/note="endopeptidase"  
/codon\_start=1  
/product="neprilysin-like peptidase beta"  
/protein\_id="AAGI8447.1"  
/db\_xref="GI:10505362"

/translation="MVRRAGWCRKKSPGVEYLMLVLLLLGAIVTLGVFYSIGLKO  
PLTSLHFWDERVVKRALRSSDICTPCVIAARILENDQSNRPCNFYOACGWLHRHV  
PETNSRYSVFDLRLEVLKGLEDTSOHPRAVERAKTLYRSQMNOVSVEKRDEPE  
LNQSVIEKRDEPELLSKLVKMGWPVMKNETMGLAKEEROLAVINSQPNRVRL  
IDLFWDDQSSRHVIIDQPTLGMSREYTFODENNHHVKARKYLEPTS VATMLRK  
DONLSKEAREMAEVLETHELANATVPQEKRHDTHLANATVPQEKRHDTHLANAT  
FNWTFLQNLSVEVELFPDEEVVGYPYILENEDIIDSARTMONYLWRVLDRIGL  
RIGLSQRFEARDVDYRKALYGTVEEYRWECVSYNMESAVGLYIKRAFSDKS  
KSTRELIERISRVFNLDDELNMMDESKKAOKAMNIROIQPYDILEDNNKH  
DEVSYLFIYDFELNGQLNKNAQRSLLKLRKVQNLIICAAVNAPYSPNRNQIV  
QIVFPAGILOPPPKDPOSLNPGFSEAFHGPRGSPMPMKRCRIW"

BASE COUNT 682 a 685 c 755 g 530 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 2652;  
Best Local Similarity 95.2%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTTCCACTTGCGGCCCATG 21  
||||||| |||||  
Db 629 AGTTCCACTTGCGGCCCATG 609

RESULT 6  
AF302077/c  
LOCUS AF302077 2694 bp mRNA linear ROD 11-JUN-2001  
DEFINITION Mus musculus neprilysin-like peptidase gamma mRNA, complete cds.  
ACCESSION AF302077  
VERSION AF302077.1 GI:10505363  
KEYWORDS SOURCE  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2694)  
Shirotsani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,  
Maruyama,K., Kiryu-Seo,S., Miyama,H., Iwata,H., Tomita,T.,  
Iwasubo,T. and Saido,T.C.  
Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most  
rapidly and efficiently among thiorphan- and phosphoramidon-sensitive endopeptidases

JOURNAL MEDLINE PUBLISHED  
REFERENCE AUTHORS TITLE JOURNAL  
Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain  
Science Institute, RIKEN, 2-1 Hirotsawa, Wako-shi, Saitama  
351-0198, Japan

```
351-0198, Japan
FEATURES
  Location/Qualifiers
    1..2694
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      25..2364
        /note="endopeptidase"
        /codon_start=1
        /product="neprilysin-like peptidase gamma"
        /protein_id="AAG18448.1"
        /db_xref="GI:10505364"
  /translation="MVERAGWCRKSPGVEYGLMVLILLGAIIVTLGVFYSIGKQL
  SSLSDICTTPSCVIAAARILENDQSRNCPENFYACGGWLRHHVIPETNSRYSVF
  IDLRDEVLKGVLEDSQHPAVEKAKTLRSCMNQSVIEKRDSPILSVLKWVG
  GWPVAMDKWNTMGLKWELETHLANATVPOEKRHDVTALYHRMDLMELQERFGLK
  LGMPRSRYFQEDNNHKYKALYEPMTSVATMLKQDLKESAMVREMAEVLLET
  HLANATVPOEKRHDVTALYHRMDLMELQERFGLKQVSLCSPGCTHSDVQAGLELG
  NPPASDRSLKGLKGFNWTFLFQNLSSVEVELFPDEEVVYGYPIYLENLEIDISYA
  RTMNYLWRLVLDIGLSQRFKEARDYRKALYGTVEEVRWRECVSYVNSNMESA
  VGSYLKRAFSDKSTVRELIEKIRSVFVNDLDELNMDESKKAKOAKAMNIREQI
  GYPDYILEDNNKHLDEEYSSLTFFEDLYFENGLOLNKNAQSLKLEKVDONLWII
  GAAVNAFVSPNRQIVFAGILOPPFESKQDQPSQINFGGIGMWIGHEITGFDNGR
  NFDKNGMLDWSFSAHFQOQSCMIYQGNFSEWELADNGVNGFTGNGIADNGR
  GVRQAYKAYLWLDGKQDLPLNTLYAQLFFINYAOWMCGSYRPEFAVQSIKTDV
  HSPLKRYVLGSLQNLPGFSEAFHCPGRGSPMHPMKRCRIW"
BASE COUNT      686 a 700 c 766 g 542 t
ORIGIN
Query Match      92.4%; Score 19.4; DB 10; Length 2694;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGTTCCTGCTGGGCCCCATG 21
      |||||
DB      560 AGTTCCTGCTGGGCCCCATG 540

RESULT 7
AF157105/c
LOCUS      AF157105      2892 bp      mRNA      linear      ROD 25-NOV-1999
DEFINITION Mus musculus soluble secreted endopeptidase mRNA, complete cds.
ACCESSION      AF157105
VERSION      AF157105.1 GI:6467398
KEYWORDS
SOURCE      Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Ikeda,K., Emoto,N., Raharjo,S.B., Nurchantari,Y., Saiki,K.,
      Yokoyama,M. and Matsuo,M.
TITLE      Molecular identification and characterization of novel
      membrane-bound metalloprotease, the soluble secreted form of which
      hydrolyzes a variety of vasoactive peptides
      J. Biol. Chem. 274 (45), 32469-32477 (1999)
JOURNAL
MEDLINE      20011457
PUBMED      10542292
REFERENCE      2 (bases 1 to 2892).
      Ikeda,K., Emoto,N. and Matsuo,M.
      Direct Submission
      Submitted (08-JUN-1999) International Center for Medical Research,
      Kobe University School of Medicine, 7-5-1 Kusunoki, Chuo, Kobe
      6500017, Japan
FEATURES
  Location/Qualifiers
    1..2892
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      281..2578
        /note="SEP; metalloprotease"
        /codon_start=1
        /product="soluble secreted endopeptidase"
        /protein_id="AAF3152.1"
        /db_xref="GI:6467399"
  /translation="MVERAGWCRKSPGVEYGLMVLILLGAIIVTLGVFYSIGKQL
  PLLTSLHFSWDETVVKRALRDSLSKSDICTTPSCVIAAARILENDQSRNCPENFY
  QYACGGWLRHHVIPETNSRYSVFIDLRDEVLKGVLEDSQHPAVEKAKTLRSCMNQSVIEKRDSPILSVLKWVG
  CMNQSVIEKRDSPILSVLKWVGHPVANDKNETMGLKWELETHLANATVPOEKRHDVTALYHRMDLMELQERFGLK
  IDLFWNDQNSRRVIYIDQPTLGMPSREYFQEDNNHKYKALYEPMTSVATMLK
  DONLSKESAMVREMAEVLLETFLANATVPOEKRHDVTALYHRMDLMELQERFGLK
  FNWTLFIQNLSSVEVELFPDEEVVYGYPIYLENLEIDISYARTMNYLWRLVLD
  RIGLSQRFKEARDYRKALYGTVEEVRWRECVSYVNSNMESAAGSLYIKRAFSDKSD
  KSTVRELIEKIRSVFVNDLDELNMDESKKAKOAKAMNIREQI GYPDYILEDNNKHL
  DEEYSSLTFFEDLYFENGLOLNKNAQSLKLEKVDONLWII GAAVNAFVSPNRN
  QIVFAGILOPPFESKQDQPSQINFGGIGMWIGHEITGFDNGRNDKNGIADNGR
  FSARHFOQSCMIYQGNFSEWELADNGVNGFTGNGIADNGVNGFTGNGIADNGR
  DGKQDLPLNTLYAQLFFINYAOWMCGSYRPEFAVQSIKTDVHSPLKRYVLGSLQNLPGFSEAFHCPGRGSPMHPMKRCRIW"
BASE COUNT      710 a 797 c 836 g 582 t
ORIGIN
Query Match      92.4%; Score 19.4; DB 6; Length 2925;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGTTCCTGCTGGGCCCCATG 21
      |||||
DB      885 AGTTCCTGCTGGGCCCCATG 865

RESULT 8
AX033272/c
LOCUS      AX033272      2925 bp      DNA      linear      PAT 21-SEP-2000
DEFINITION Sequence 12 from Patent WO0047750.
ACCESSION      AX033272
VERSION      AX033272.1 GI:10280087
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      Boileau,G. and Desgroselliers,L.
      1 (bases 1 to 2925)
TITLE      New metalloproteases of the neprilysin family
      Patent: WO 0047750-A 12 17-AUG-2000;
      BOILEAU GUY (CA); DESGROSSELLERS LUC (CA); UNIVERSITE DE MONTREAL
      (CA)
FEATURES
  Location/Qualifiers
    1..2925
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      332..2629
        /note="unnamed protein product"
        /codon_start=1
        /protein_id="CAC09977.1"
        /db_xref="GI:10280088"
  /translation="MVERAGWCRKSPGVEYGLMVLILLGAIIVTLGVFYSIGKQL
  PLLTSLHFSWDETVVKRALRDSLSKSDICTTPSCVIAAARILENDQSRNCPENFY
  QYACGGWLRHHVIPETNSRYSVFIDLRDEVLKGVLEDSQHPAVEKAKTLRSCMNQSVIEKRDSPILSVLKWVG
  CMNQSVIEKRDSPILSVLKWVGHPVANDKNETMGLKWELETHLANATVPOEKRHDVTALYHRMDLMELQERFGLK
  IDLFWNDQNSRRVIYIDQPTLGMPSREYFQEDNNHKYKALYEPMTSVATMLK
  DONLSKESAMVREMAEVLLETFLANATVPOEKRHDVTALYHRMDLMELQERFGLK
  FNWTLFIQNLSSVEVELFPDEEVVYGYPIYLENLEIDISYARTMNYLWRLVLD
  RIGLSQRFKEARDYRKALYGTVEEVRWRECVSYVNSNMESAAGSLYIKRAFSDKSD
  KSTVRELIEKIRSVFVNDLDELNMDESKKAKOAKAMNIREQI GYPDYILEDNNKHL
  DEEYSSLTFFEDLYFENGLOLNKNAQSLKLEKVDONLWII GAAVNAFVSPNRN
  QIVFAGILOPPFESKQDQPSQINFGGIGMWIGHEITGFDNGRNDKNGIADNGR
  FSARHFOQSCMIYQGNFSEWELADNGVNGFTGNGIADNGVNGFTGNGIADNGR
  DGKQDLPLNTLYAQLFFINYAOWMCGSYRPEFAVQSIKTDVHSPLKRYVLGSLQNLPGFSEAFHCPGRGSPMHPMKRCRIW"
BASE COUNT      710 a 797 c 836 g 582 t
ORIGIN
Query Match      92.4%; Score 19.4; DB 6; Length 2925;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 AGTTCACCTGGGCCCATG 21
|||||
Db 936 AGTTCACCTGGGCCCATG 916

RESULT 9
LOCUS AF176569/c
DEFINITION Mus musculus neprilysin-like metalloproteinase 1 (N11) mRNA.
ACCESSION AF176569
VERSION AF176569.1 GI:7769082
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 2925)
AUTHORS Ghaddar,G., Ruchon,A.F., DesGroselliers,L. and Boileau,G.
TITLE Molecular cloning and biochemical characterization of a new mouse testis soluble-zinc-metalloproteinase of the neprilysin family
JOURNAL Biochem. J. 347 (Pt 2), 419-429 (2000)
MEDLINE 21060448
PUBMED 10749671
REFERENCE 2 (bases 1 to 2925)
AUTHORS Ghaddar,G., Ruchon,A.F., DesGroselliers,L. and Boileau,G.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1999) Biochemistry, University of Montreal, 2900 Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES
source
1. .2925
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="testis"
1. .2925
/gene="N11"
332. .2629
/gene="N11"
/codon_start=1
/product="neprilysin-like metalloproteinase 1"
/db_xref="GI:7769083"
/translation="MVFRAGWCKKSPGFVEYGLMVLVLLLLGAVTLGVFYSGIKOL
PLITSLHFSWDETVKRALRDSLSKSDICTPSCVIAARILENMDQSRNCFNY
QVACGGLRHVYTPETNSRYSVFDILRDEVLKGVLEDSTSOHRPAVEKAKTLVRS
CMNQSVEIKRDSEPLSLVKMGWVPMVKWNETGLKWELEQLAVLANSQFNRRVL
IDLFIWNDQNSRHVYIDQPLGHPSEHYEQEDNNHKVRKAYLEFWTSVATMLRK
DONLSKMAVREEMAEVLELETHLANATVPQERHDVTALYHRMDLMEIQERVLGK
FNWTLFIQNLVSVEVEFPDEEVYVYIPLENLHDIIDSYGARTMONTLVRLVLD
RIGLSQREKARVDYRKALYGTVEVRWRECVSYVNSNNEAVGSLYIKRAFSDKS
KSTVRELKIRSVFVDNLDELNMWDEESKKKQAEKAMNIREOIGYDPDYILEDNKHL
DEYSLSITYEDLYFENGLQNLNNAQRSLLKLRKVDQNLNLIIGAVNVAFSNRRN
QVYFPAIGLIPPPSKDQPSLNFGGIVGHIETHFGDDNGRDNKMDLWNSN
FSARHFQDSQCMYIQGNFSEWELADQNVNGSTLGENIADNGGVQAOYKAYLRWLA
DGKDLQRLNLNTYLAQFIINTYAOVWCGSYRPEFAVQSIKTVDVHSLPKRYVLGSLQ
LPQFSFAFHCPRGSPMHPMKRCRIW"
BASE COUNT 710 a 797 c 836 g 582 t
ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 2925;
Best Local Similarity 95.2%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTTCACCTGGGCCCATG 21
|||||
Db 936 AGTTCACCTGGGCCCATG 916

RESULT 10
LOCUS AE000795/c
DEFINITION Methanobacterium thermoautotrophicum str. Delta H.
Methanothermobacter thermoautotrophicus str. Delta H
Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
Methanobacteriaceae; Methanothermobacter.
REFERENCE 1 (bases 1 to 10208)
AUTHORS Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H., Dubois,J.,
Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R., Gilbert,K.,
Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothier,B., Qiu,D.,
Spadafora,R., Vicaire,R., Wang,Y., Wierzbowski,J., Gibson,R.,
Jiwani,N., Caruso,A., Bush,D. and Reeve,J.N.
TITLE Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics
JOURNAL J. Bacteriol. 179 (22), 7135-7155 (1997)
MEDLINE 98037514
PUBMED 9371463
REFERENCE 2 (bases 1 to 10208)
AUTHORS Smith,D.R.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1997) Genomics and Technology Development, Genome
Therapeutics Corporation, 100 Beaver Street, Waltham, MA
02154-8448, USA
FEATURES
source
1. .10208
/organism="Methanothermobacter thermoautotrophicus str.
Delta H"
/strain="Delta H"
/db_xref="taxon:187420"
/clone="MTH"
/notes="synonym: Methanobacterium thermoautotrophicum str.
deltaH"
171. .977
/gene="MTH1"
171. .977
/gene="MTH1"
/notes="Function Code:14.01 - Unknown, Conserved protein;
similar to, sp:LN:YL31_HALMA AC:P20571, p(1)=1.2E-16,
pid=278"
/codon_start=1
/trans_table=11
/product="conserved protein"
/protein_id="AAB84510.1"
/db_xref="GI:2621037"
/translation="MNRVDSLIFIPDSLTAEGLKIKTYKVLIAAASIFGVKRV
IYHDADGEARFIRDLITYMDTPQYLRKKVPIRELKHVGLPLPLPHPTGKPV
GEYQGLTVKRVKGLTVDIGADKALCKRLKLVNRIMSFRVRLKGLKEILPEDEP
RYNGEYLDTRNLAEKLTGVADVVATSNASPTISILDEVKTRMRGAAALFG
GPYKGLPEIDAIWVNTLPGQCTETVTEEAVALTSVFNMLTQIDERKE"
997. .2010
/gene="MTH2"
997. .2010
/gene="MTH2"
/notes="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RL3_METJA AC:P54014,
p(1)=2.1E-95, pid=538"
/codon_start=1
/trans_table=11
/product="ribosomal protein L3 (E.coli L3)"
/protein_id="AAB84521.1"
/db_xref="GI:2621048"
/translation="MARHOPKGVSAFSPRKAARETPRVKSWPQVDPGGLALAGY
KAGTHYMWVNDNKNSTEGMEVSTPVLLEVPPLTVAVRYEYKTLGLGLVLA
TETKDDLRRLKLPADDYDQAAIEKIRSNMEYVADRVYVHINPLASVPKKKEPV
ECGLGKTPKEKFEYALEILGKDRASIFSEGAFVDAIVTGFQFGQPVKRWGIRI
QYKAARHSKGLSGPWPSPRTMTVPQAGMGYHRRTEYNKQILKIGDASDL
VNPDGFFRYGLVRNDYVMIKSGVPGPKRLVLRKRAAGKQEAPOINYSTASK
QGV"
2023. .2787
/gene="MTH3"
```

```

CDS      2023..2787
/ gene="MTH3"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RL4_METJA AC:P54015,
p()=1.8E-68, pid=55%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein L4 (E.coli L4)"
/ protein_id="AAB84523.1"
/ db_xref="GI:2621050"
/ translation="MKIKVYSLGEAIDEMELPEIFNEERPDPVIKRAVLSAQARVQ
PWGPDPMAGRTSQAGAGVAMPRKNSRAAFVQAVGRRRAHPPRPKNYHE
RINKERLARSAAATARKDLVEARGHRIENVPLVVDDELSTMKRTADTRVFF
KILGIMDDIVAREGKIKRAGKMGKRYTTPRGPLTVVGDDKGTGRGARNHPGVDV
VRVENLNAELLAPGTHPGRLTVFTRSAIEKLDELFO"
/ gene="MTH4"
/ gene="MTH4"
/ gene="MTH4"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RL23_METJA
AC:P54016, p()=3.3E-19, pid=41%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein L23a (E.coli L23)"
/ protein_id="AAB84524.1"
/ db_xref="GI:2621051"
/ translation="MMNYSKSGVTVYIMDPYAVIMKPHVTEKSMNLIDQNNEAFVY
MRKSTKDVRAFEELFAVKVERVNTVTPRGOKIAYIKLAKEHSAEDIAVKLGVF"
/ gene="MTH5"
/ gene="MTH5"
/ gene="MTH5"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RL2_METJA AC:P21479,
p()=2.1E-79, pid=56%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein L8 (E.coli L2)"
/ protein_id="AAB84525.1"
/ db_xref="GI:2621052"
/ translation="MKRLISQRRGRTPTYSASHRFKGIKRYAYDSTESGSLKG
KVVDIHPDGPAPVAVKVFENGEEELIAPALMLNEECVGKARVPGNSLPSE
IPGTIYNIENRPGDGLVRSSTGYASLIHDADKAVIELPSGELKALNPQCRATV
GVVAGGRREKPPKAGKYHALRAKGSVTVRGVAMNAVDPHGGNGNQHPGRPTT
VSRHAPPGRKVGSIARRTGKR"
/ gene="MTH6"
/ gene="MTH6"
/ gene="MTH6"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RS19_METJA
AC:P54018, p()=2.1E-40, pid=57%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein S15 (E.coli S19)"
/ protein_id="AAB84526.1"
/ db_xref="GI:2621053"
/ translation="MARKEFRVGYTLELQEMPLDDVVKLFPSRRRSRLKRGFLPRQ
KVLKIRIKKEGKTEGPPVIRTHCRDMIVLPENVMGTFGIHNGKEFVVKIQPEM
ICGYGFEAPTRKKVEHGDGGMGATRFSMFVLK"
/ gene="MTH7"
/ gene="MTH7"
/ gene="MTH7"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RL22_METJA
AC:P54033, p()=1.4E-41, pid=55%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein L17 (E.coli L22)"
/ protein_id="AAB84527.1"
/ db_xref="GI:2621054"
/ translation="MAKVYAYKEEDRSTARSATHLKI5PKHVAEICREIRGMELE
KAKYLVIEVIRMEPVAKRYNRYKHGRGLNGWASGRYPVKAAGQILKVLNAAENA
EYKGLDTEKLRIIHISSHRGPVIRGWIAPFRAFPFNTPTTHVQIVLGEA"
/ gene="MTH8"
/ gene="MTH8"
/ gene="MTH8"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RS3_METJA AC:P54034,
p()=8.5E-53, pid=39%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein S3 (E.coli S3)"
/ protein_id="AAB84528.1"
/ db_xref="GI:2621055"
/ translation="MIEKDFVVEGLRRTTRIDYLEKELEKLAGYGGMDVQVTPMGTMVY
VYRPGMVGIRGGKTVRATIKLKNKFDLPQVEVKEVDVPELNPKIHAHIAAML
QYGHFRFVAYTMMRIIMAGAGQGVETISGIRGARSATATFTGYIKKCGEPSVKH
VRGCFATVQLKPGVLGVYVIRMPDPVLPDKVEIDPRVTETPAPAESEASEVDELE
EVEDELEDELEVEDELEVEDELEDETEAEKDDAGESEK"
complement(5362..5396)
/ note="35 bp direct repeat includes part of MTH8
(ribosomal protein S3); 97% ID to interval 5450-5416"
complement(5416..5450)
/ note="35 bp direct repeat includes part of MTH8
(ribosomal protein S3); 97% ID to interval 5396-5362"
5546..5740
/ gene="MTH9"
/ gene="MTH9"
/ gene="MTH9"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:RL29_METJA
AC:P54035, p()=4.5E-15, pid=58%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein L35 (E.coli L29)"
/ protein_id="AAB84529.1"
/ db_xref="GI:2621056"
/ translation="MAILRSEIREMDGELOKLDELKAEYARYISKAAAGIHENP
GMREIRRTIARVLTIMNEK"
5753..6058
/ gene="MTH10"
/ gene="MTH10"
/ gene="MTH10"
/ note="Function Code:10.04 - Metabolism of Macromolecules,
Ribosomal proteins; similar to, sp:LN:YRP1_METVA
AC:P14021, p()=4.8E-27, pid=55%"
/ codon_start=1
/ transl_table=11
/ product="ribosomal protein SUI1"
/ protein_id="AAB84511.1"
/ db_xref="GI:2621038"
/ translation="MKICDVGCLPEELVCEIAREVQTLKVYVRRRFGKVMVTIEG
Query Match      87.68; Score 18.4; DB 1; Length 10208;
Best Local Similarity 95.08; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGTTCCCACTTGGGCGCCAT 20
      |||||
Db      9352 AGTTCCCACTTGGGCGCCTT 9333

RESULT 11
D42021S07/c
LOCUS      D42021S07      492 bp      DNA      linear      MAM 14-APR-2000
DEFINITION      Canis familiaris BGT1 gene, exon 6.
ACCESSION      D42027
VERSION      D42027.1      GI:2463595
KEYWORDS      alternative splicing.
SEGMENT      7 of 17
SOURCE      Canis familiaris DNA.
ORGANISM      Canis familiaris

```



```

repeat_region      /rpt_family="Alu"
7703..7989
repeat_region      /rpt_family="Alu"
9071..9348
repeat_region      /rpt_family="Alu"
complement(12236..12528)
repeat_region      /rpt_family="MER39"
complement(12272..12390)
repeat_region      /rpt_family="MER21"
complement(13571..13775)
repeat_region      /rpt_family="L1"
17618..17913
repeat_region      /rpt_family="Alu"
17980..18241
repeat_region      /rpt_family="Alu"
18387..18681
repeat_region      /rpt_family="Alu"
complement(19415..19479)
repeat_region      /rpt_family="MIR"
19591..19878
repeat_region      /rpt_family="Alu"
20022..20189
repeat_region      /rpt_family="MER20"
21380..21523
repeat_region      /rpt_family="MIR"
complement(21888..22093)
repeat_region      /rpt_family="MER33"
complement(22100..22237)
repeat_region      /rpt_family="Alu"
22532..22705
repeat_region      /rpt_family="MIR"
24353..24599
repeat_region      /rpt_family="Alu"
24868..25009
repeat_region      /rpt_family="MIR"
25152..25297
repeat_region      /rpt_family="Alu"
26324..28673
repeat_region      /rpt_family="L1"
complement(27190..27736)
repeat_region      /rpt_family="MER2B2"
30834..31092
repeat_region      /rpt_family="Alu"
31710..31981
repeat_region      /rpt_family="Alu"
32871..33161
repeat_region      /rpt_family="Alu"
33719..34095
repeat_region      /rpt_family="L1"
complement(34126..34424)
repeat_region      /rpt_family="Alu"
complement(34528..34854)
repeat_region      /rpt_family="Alu"

BASE COUNT 10612 a 7769 c 7629 g 8854 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 9; Length 34864;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTCCCACTGGGGCCCATG 21
||||||| 1 |||||||
Db 2057 AGTCCCACTGGGGCCCATG 2077

RESULT 14
CNS08C8T
LOCUS      CNS08C8T      84181 bp      DNA      linear      HTG 10-JUL-2002
DEFINITION Oryza sativa chromosome 12 clone OJ1122_G07.*** SEQUENCING IN
PROGRESS ***; 3 ordered pieces.
ACCESSION  AL731889
VERSION     AL731889.2      GI:21732199
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.

SOURCE      Oryza sativa.
ORGANISM    Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 84181)
Choi S.N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P.,
Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,
Weissenbach, J., and Quetier, F.
Oryza sativa chromosome 12 sequencing
Unpublished
2 (bases 1 to 84181)
Genoscope.
Direct Submission
Submitted (09-JUL-2002) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Jul 11, 2002 this sequence version replaced gi:20372839.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence.
Work on the sequence is in progress and the release of this data is
based on the understanding that the sequence may change as work
continue. The sequence may be contaminated with foreign sequence
from E.coli, yeast, vector, phage, etc.
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and Genoscope sequencing data.
Contigs composition :
53931 bp contig from 1 to 53931
17812 bp contig from 54032 to 71843
12238 bp contig from 71944 to 84181.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 53931: contig of 53931 bp in length
* 53932 54031: gap of 100 bp
* 54032 71843: contig of 17812 bp in length
* 71844 71943: gap of 100 bp
* 71944 84181: contig of 12238 bp in length.
FEATURES
Location/Qualifiers
1..84181
/organism="Oryza sativa"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:4530"
/chromosome="12"
/clone="OJ1122_G07"
/clone_lib="Monsanto"
BASE COUNT 22856 a 19004 c 18666 g 23455 t 200 others
ORIGIN
Query Match      84.8%; Score 17.8; DB 2; Length 84181;
Best Local Similarity 90.5%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches -2; Indels 0; Gaps 0;

QY 1 AGTCCCACTGGGGCCCATG 21
||||||| 1 |||||||
Db 80214 AGTCCCACTGGGGCCCATG 80234

RESULT 15
CNS08C8V/c
LOCUS      CNS08C8V      99239 bp      DNA      linear      HTG 29-APR-2002
DEFINITION Oryza sativa chromosome 12 clone Monsanto-OJ1374_A04, ***
SEQUENCING IN PROGRESS ***; 5 ordered pieces.
ACCESSION  AL731891
VERSION     AL731891.1      GI:20372841
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.

```

SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE 1 (bases 1 to 99239)  
AUTHORS Choisine.N., Orjeda.G., Cattolico.L., Demange.N., Wincker.P.,  
Seguren.B., Pellotier.E., Scarpelli.C., Salanoubat.M.,  
Weissenbach.J. and Quetier.F.  
TITLE Oryza sativa chromosome 12 sequencing  
JOURNAL Unpublished  
AUTHORS 2. (bases 1 to 99239)  
GENOSCOPE.  
Direct Submission  
Submitted (29-APR-2002) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segrif@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence.  
Work on the sequence is in progress and the release of this data is  
based on the understanding that the sequence may change as work  
continue. The sequence may be contaminated with foreign sequence  
from E.coli, yeast, vector, phage, etc.  
The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and Genoscope sequencing data.  
Contigs composition :  
11584 bp contig from 1 to 11584  
36246 bp contig from 11685 to 47930  
39114 bp contig from 48031 to 87144  
9829 bp contig from 87245 to 97073  
2066 bp contig from 97174 to 99239.  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 5 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 11584: contig of 11584 bp in length  
\* 11585 11684: gap of 100 bp  
\* 11685 47930: contig of 36246 bp in length  
\* 47931 48030: gap of 100 bp  
\* 48031 87144: contig of 39114 bp in length  
\* 87145 87244: gap of 100 bp  
\* 87245 97073: contig of 9829 bp in length  
\* 97074 97173: gap of 100 bp  
\* 97174 99239: contig of 2066 bp in length.  
FEATURES  
source Location/Qualifiers  
1. .99239  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="12"  
/clone="Monsanto-OJ1374\_A04"  
/clone\_lib="Monsanto"  
BASE COUNT 28380 a 21094 c 21377 g 27988 t 400 others  
ORIGIN  
Query Match 84.8%; Score 17.8; DB 2; Length 99239;  
Best Local Similarity 90.5%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 AGTCCACATGGGGCCCATG 21  
||| ||||| ||||| |||||  
Db 97242 AGCTCCCTCTGGGGCCCATG 97222

Search completed: July 8, 2003, 03:34:22  
Job time : 236.102 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 131.941 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-6

Perfect score: 21

Sequence: 1 agttccacttggggccatg 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq\_101002.\*  
1: /SID22/qcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/qcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/qcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/qcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/qcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/qcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/qcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/qcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/qcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/qcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/qcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/qcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/qcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/qcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/qcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/qcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/qcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/qcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/qcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/qcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/qcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/qcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/qcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/qcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	20	AZ28812
2	21	100.0	2765	20	AZ28810
3	19.4	92.4	2286	24	AAD28130
4	19.4	92.4	2925	21	AAAG3763
5	17.8	84.8	520	21	AAC41442
6	16.8	80.0	2026	24	AB072543
7	16.8	80.0	2235	24	ABL64430
8	16.8	80.0	2304	21	AAC74249
9	16.8	80.0	2487	22	AAI58283

c 10	16.8	80.0	2745	23	ABK43576	DNA encoding novel
c 11	16.8	80.0	3443	22	RAI60069	Human polynucleoti
c 12	16.8	80.0	5762	22	AAF54792	Nucleotide sequenc
c 13	16.8	80.0	6607	22	AAF54812	Nucleotide sequenc
c 14	16.8	80.0	14044	22	AAF54793	Nucleotide sequenc
c 15	16.8	80.0	23024	22	AAF25499	Nucleotide sequenc
c 16	16.4	78.1	2635	22	AA41164	CDNA encoding novel
c 17	16.4	78.1	2635	23	ABK43877	DNA encoding novel
c 18	16.4	78.1	2775	21	AA93875	Human beta-1,3 gal
c 19	16.4	78.1	10562	21	AAA93876	Human beta3gal-T5
c 20	16.4	78.1	12280	22	ABAI9083	Human nervous syst
c 21	16.4	78.1	12280	22	ABK76794	Human immune/haema
c 22	16.2	77.1	402	22	ABA57025	Human foetal liver
c 23	16.2	77.1	476	21	ARC41451	Zea mays DNA fragm
c 24	16.2	77.1	549	22	AAK61369	Human immune/haema
c 25	16.2	77.1	555	22	AAK79931	Human immune/haema
c 26	16.2	77.1	555	22	AAK79932	Human immune/haema
c 27	16.2	77.1	555	22	AAK79933	Human immune/haema
c 28	16.2	77.1	1080	23	ABL10743	Drosophila melanog
c 29	16.2	77.1	1298	22	AS00835	Human CDNA clone H
c 30	16.2	77.1	2013	22	AA04910	Human secreted pro
c 31	16.2	77.1	2127	22	AAF74440	Human PRO9 nucleot
c 32	16.2	77.1	2127	22	AAF74441	Human PRO10 nucle
c 33	16.2	77.1	2143	22	AAF74443	Human PRO12 nucle
c 34	16.2	77.1	3316	23	ABL10742	Drosophila melanog
c 35	16.2	77.1	15848	20	AZ332190	Human heparin cofa
c 36	16.2	77.1	15849	24	ABN95864	Gene #2362 used to
c 37	16	76.2	812	22	AAH08380	Human CDNA clone (
c 38	15.8	75.2	437	20	AAK16713	Mutant mouse tub g
c 39	15.8	75.2	437	20	AAK16714	Wild type mouse tu
c 40	15.8	75.2	480	18	AAK96638	Mouse tub mutation
c 41	15.8	75.2	480	21	AAK94631	Partial mutant mou
c 42	15.8	75.2	519	22	AAK68995	Human immune/haema
c 43	15.8	75.2	522	22	AAK56857	Human immune/haema
c 44	15.8	75.2	522	22	AAK68996	Human immune/haema
c 45	15.8	75.2	522	22	AAK68997	Human immune/haema

ALIGNMENTS

RESULT 1  
AAZ28812  
ID AAZ28812 standard; DNA; 21 BP.  
XX AC AAZ28812;  
XX DT 01-FEB-2000 (first entry)  
XX DE Rat membrane metalloprotease NEPII gene probe #2.  
XX KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
XX OS Synthetic.  
XX OS Rattus rattus.  
XX PN FR2777291-Al.  
XX PD 15-OCT-1999.  
XX PF 08-APR-1998; 98PR-0004389.  
XX PR 08-APR-1998; 98PR-0004389.  
XX PA (INRM ) INSM INST NAT SANTE & RECH MEDICALE.  
XX PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
XX PI Schwartz JC;  
XX WPI; 1999-593429/51.  
DR

XX New membrane metalloprotease NEPII, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 3; Page 20; 29pp; French.  
 XX  
 CC Sequences AA28811-428827 represent probes for detecting the rat  
 CC membrane metalloprotease designated neprilysine II (NEPII) gene  
 CC (AA28810). NEPII is involved in (in)activation of neuronal and hormonal  
 CC peptide messengers. NEPII is used to screen for specific substrates (used  
 CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
 CC used to detect NEPII or for treatment of disorders related to peptidergic  
 CC signalling in which NEPII is involved, e.g. cardiovascular or  
 CC neurodegenerative diseases; growth disorders of endocrine origin;  
 CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
 CC disorders.  
 XX  
 SQ Sequence 21 BP; 3 A; 7 C; 6 G; 5 T; 0 other;  
 XX  
 Query Match 100.0%; Score 21; DB 20; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 AGTCCCACTGGGGCCCATG 21  
 |||||  
 DB 1 AGTCCCACTGGGGCCCATG 21  
 |||||  
 RESULT 2  
 AA28810/c  
 ID AA28810 standard; CDNA; 2765 BP.  
 XX  
 AC AA28810;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Rat membrane metalloprotease NEPII gene.  
 XX  
 KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
 KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
 KW cardiovascular disease; neurodegenerative disease; growth disorder;  
 KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
 XX  
 OS Rattus rattus.  
 XX  
 PN FR2777291-A1.  
 XX  
 PD 15-OCT-1999.  
 XX  
 PF 08-APR-1998; 98FR-0004389.  
 XX  
 PR 08-APR-1998; 98FR-0004389.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
 PI Schwartz JC;  
 XX  
 WPI: 1999-593429/51.  
 DR P-PSDB; AAY44177.  
 DR  
 XX New membrane metalloprotease NEPII, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 2; Page 12-16; 29pp; French.  
 XX  
 CC This sequence represents the gene for the rat membrane metalloprotease  
 CC designated neprilysine II (NEPII), which is involved in (in)activation  
 CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
 CC for specific substrates (used to detect NEPII in cells and tissues) or  
 CC inhibitors, which can also be used to detect NEPII or for treatment of

CC disorders related to peptidergic signalling in which NEPII is involved,  
 CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
 CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
 CC endocrine disorders.  
 XX  
 SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
 XX  
 Query Match 100.0%; Score 21; DB 20; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 AGTCCCACTGGGGCCCATG 21  
 |||||  
 DB 741 AGTCCCACTGGGGCCCATG 721  
 |||||  
 RESULT 3  
 AAD28130/c  
 ID AAD28130 standard; DNA; 2286 BP.  
 XX  
 AC AAD28130;  
 XX  
 DT 07-MAY-2002 (first entry)  
 XX  
 DE Soluble secreted endopeptidase (SEP) consensus DNA.  
 XX  
 KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
 KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
 KW FSD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
 XX  
 OS Homo sapiens.  
 OS Mus sp.  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1664..2286  
 FT /tag= a  
 FT /note= "Encodes catalytic domain"  
 XX  
 PN WO200206492-A1.  
 XX  
 PD 24-JAN-2002.  
 XX  
 PF 16-JUL-2001; 2001WO-IB01263.  
 XX  
 PR 14-JUL-2000; 2000GB-0017387.  
 XX  
 PA (PFIZ ) PFIZER LTD.  
 PA (PFIZ ) PFIZER INC.  
 XX  
 PI Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
 XX  
 WPI: 2002-155042/20.  
 DR  
 XX An isolated and/or purified nucleic acid encoding a human soluble  
 PT secreted endopeptidase which is useful for treating sexual dysfunction,  
 PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
 PT female sexual arousal disorder -  
 XX  
 PS Disclosure; Fig 6; 167pp; English.  
 XX  
 CC The invention relates to an isolated and/or purified nucleic acid  
 CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
 CC antibody and the compound which inhibits or selectively inhibits the  
 CC human SEP protein are useful in the manufacture of a medicament for  
 CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
 CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
 CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
 CC also useful for treating the above disorders and other disorders such  
 CC as premature ejaculation; anorgasmia, sexual pain (e.g. vaginismus) and  
 CC hypoaffective sexual desire disorder. The present sequence is SEP consensus  
 CC DNA sequence found in human, mouse and rat.

XX SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
Query Match 92.4%; Score 19.4; DB 24; Length 2286;  
Best Local Similarity 95.2%; Pred. No. 8.2;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTTCACACTGGGCCCATG 21  
|||||  
Db 617 AGTTCACACTGGGCCCATG 597

RESULT 4  
AAAG3763/c  
ID AAAG3763 standard; CDNA; 2925 BP.  
XX AC  
XX AAAG3763;  
DT 04-DEC-2000 (first entry)  
XX DE  
XX cDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.  
XX KW Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
KW NEP-like enzyme; protein production; protein secretion;  
KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
KW fertility; bone disease; abnormal phosphate metabolism; ss.  
XX OS Mus sp.  
XX FH Key Location/Qualifiers  
XX CDS 332..2629  
XX FT /\*tag= a  
XX FT /product= "neutral endopeptidase metalloproteinase-like  
XX enzyme NL-1"  
XX PN W0200047750-A2.  
XX PD 17-AUG-2000.  
XX PF 11-FEB-2000; 2000WO-CA00147.  
XX PR 11-FEB-1999; 99CA-2260376.  
XX PA (UYMO-) UNIV MONTREAL.  
XX PI Desgroseillers L, Boileau G;  
XX DR WPI: 2000-549148/50.  
XX DR P-PSDB; AAB08130.  
XX PT Novel neutral endopeptidase-like metalloproteinase polypeptides and  
XX polynucleotides, used to screen for related sequences and enzyme  
XX inhibitors, used for the treatment of NL-3 related bone disorders -  
XX Disclosure; Fig 3; 59pp; English.  
XX CC The present sequence encodes a murine neutral endopeptidase  
XX metalloproteinase-like enzyme, designated NL-1. The specification  
XX also describes NL-2 and NL-3. The NL enzymes are used to test for  
XX specific inhibitors. The N-terminal region of the enzymes can be used  
XX to promote production and secretion of foreign proteins and active  
XX biopeptides, using chimeric constructs containing the foreign protein  
XX downstream from and in phase with the N-terminal region. The NL enzymes  
XX are have been localised to the brain, and may be useful in the  
XX treatment of neurological diseases such as Alzheimer's disease, pain,  
XX and psychiatric disorders. NL enzymes have also been localised to the  
XX testis and ovaries, and may be used to control fertility. They have  
XX also been localised to bones, and may be used to treat bone diseases,  
XX and abnormal phosphate metabolisms related to improper peptide  
XX processing by the NL-3 enzyme.  
XX SQ Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;

Query Match 92.4%; Score 19.4; DB 21; Length 2925;  
Best Local Similarity 95.2%; Pred. No. 8.3;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTTCACACTGGGCCCATG 21  
|||||  
Db 936 AGTTCACACTGGGCCCATG 916

RESULT 5  
AAC41442/c  
ID AAC41442 standard; DNA; 520 BP.  
XX AC  
XX AAC41442;  
DT 17-OCT-2000 (first entry)  
XX DE  
XX Zea mays DNA fragment SEQ ID NO: 31897.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic;  
KW pathway; promoter; termination sequence; corn; ss.  
XX OS Zea mays subsp. mays.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0132484.  
XX PR 05-MAY-1999; 99US-0132485.  
XX PR 06-MAY-1999; 99US-0132486.  
XX PR 06-MAY-1999; 99US-0132487.  
XX PR 07-MAY-1999; 99US-0132863.  
XX PR 11-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.  
XX PR 14-MAY-1999; 99US-0134219.  
XX PR 14-MAY-1999; 99US-0134221.  
XX PR 14-MAY-1999; 99US-0134370.  
XX PR 18-MAY-1999; 99US-0134768.  
XX PR 19-MAY-1999; 99US-0134941.  
XX PR 20-MAY-1999; 99US-0135124.  
XX PR 21-MAY-1999; 99US-0135353.  
XX PR 24-MAY-1999; 99US-0135629.  
XX PR 25-MAY-1999; 99US-0136021.  
XX PR 27-MAY-1999; 99US-0136392.  
XX PR 28-MAY-1999; 99US-0136782.  
XX PR 01-JUN-1999; 99US-0137222.  
XX PR 03-JUN-1999; 99US-0137528.  
XX PR 04-JUN-1999; 99US-0137502.  
XX PR 07-JUN-1999; 99US-0137724.  
XX PR 08-JUN-1999; 99US-0138094.  
XX PR 10-JUN-1999; 99US-0138540.  
XX PR 10-JUN-1999; 99US-0138847.



## RESULT 6

ABQ72543/c

ID ABQ72543 standard; cDNA; 2026 BP.

XX AC ABQ72543;

XX DT 03-SEP-2002 (first entry)

XX DE Human MDDT encoding cDNA SEQ ID NO 95.

XX Human; MDDT; disease detection and treatment molecule polynucleotide;  
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
 KW rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;  
 KW hepatocytic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
 KW antiallergic; antianemic; antiasthmatic; antiatherosclerotic; antigout;  
 KW neuroprotective; antirheumatic; antiarthritic; gene; ss.  
 XX OS Homo sapiens.

XX PN W0200240715-A2.

XX PD 23-MAY-2002.

XX PF 06-SEP-2001; 2001WO-US27628.

XX PR 06-SEP-2000; 2000US-230505P.

XX PR 06-SEP-2000; 2000US-230514P.

XX PR 06-SEP-2000; 2000US-230515P.

XX PR 06-SEP-2000; 2000US-230517P.

XX PR 06-SEP-2000; 2000US-230518P.

XX PR 06-SEP-2000; 2000US-230519P.

XX PR 06-SEP-2000; 2000US-230597P.

XX PR 06-SEP-2000; 2000US-230598P.

XX PR 06-SEP-2000; 2000US-230599P.

XX PR 06-SEP-2000; 2000US-230610P.

XX PR 06-SEP-2000; 2000US-230865P.

XX PR 06-SEP-2000; 2000US-230988P.

XX PR 06-SEP-2000; 2000US-230989P.

XX PR 07-SEP-2000; 2000US-230951P.

XX PR 07-SEP-2000; 2000US-231163P.

XX PR 07-SEP-2000; 2000US-231167P.

XX PA (INCY-).INCYTE GENOMICS INC.

XX PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;

XX PI Jones AL, Yu JY, Wright RJ, Gletzen D, Liu TF, Yap PE, Dahl CR;

XX PI Momiya MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;

XX PI Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V, Daiffo A;

XX PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;

XX DR WPI: 2002-527544/56.

XX DR P-PSDB; ABP51325.

XX PT Novel human disease detection and treatment polypeptide, useful in

XX PT diagnosis, prevention or treatment of cell proliferative disorders e.g.

XX PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder

XX PT e.g. AIDS

XX PS Claim 1; Page 361-362; 618pp; English.

XX CC The invention relates to an isolated human disease detection and

XX CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a

XX CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the

XX CC specification, a naturally occurring polypeptide comprising a sequence

XX CC having at least 90% identity to (I) or a biologically active or

XX CC immunogenic fragment of (I). (I) is useful for screening a compound for

XX CC effectiveness as an agonist or antagonist, for screening a compound that

XX CC specifically binds (I) or modulates the activity of (I), and for

XX CC preparing a polyclonal or monoclonal antibody by hybridoma technology.

CC Nucleic acids (II) (AB072449-AB072700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDDT.  
 CC (I) or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDDT, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germ-line gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences.

XX SQ Sequence 2026 BP; 296 A; 631 C; 742 G; 357 T; 0 other;

XX Query Match 80.0%; Score 16.8; DB 24; Length 2026;

XX Best Local Similarity 90.0%; Pred. No. 1.4e+02;

XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGTTCCTCCACTGGGGCCCAT 20

Db 789 AGTGGCCACTGGAGCCCAT 770

RESULT 7

ABL64430/c

ID ABL64430 standard; DNA; 2235 BP.

XX AC ABL64430;

XX DT 15-MAY-2002 (first entry)

XX DE Stomach cancer related gene sequence SEQ ID NO:2767.

XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

XX KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;

XX KW gene; ds.

XX OS Homo sapiens.

XX PN WO200194629-A2.

XX PD 13-DEC-2001.

XX PF 30-MAY-2001; 2001WO-US10838.

XX PR 05-JUN-2000; 2000US-209473P.

XX PR 05-JUN-2000; 2000US-209531P.

XX PR 18-SEP-2000; 2000US-233133P.

XX PR 18-SEP-2000; 2000US-233617P.

XX PR 20-SEP-2000; 2000US-234009P.

XX PR 20-SEP-2000; 2000US-234034P.

XX PR 20-SEP-2000; 2000US-234052P.

XX PR 22-SEP-2000; 2000US-234509P.

XX PR 22-SEP-2000; 2000US-234567P.

XX PR 25-SEP-2000; 2000US-234923P.

XX PR 25-SEP-2000; 2000US-234924P.

XX PR 25-SEP-2000; 2000US-235077P.

XX PR 25-SEP-2000; 2000US-235082P.

XX PR 25-SEP-2000; 2000US-235134P.

XX PR 25-SEP-2000; 2000US-235280P.

XX PR 26-SEP-2000; 2000US-235637P.

PR 26-SEP-2000; 2000US-235638P.  
 PR 27-SEP-2000; 2000US-235711P.  
 PR 27-SEP-2000; 2000US-235720P.  
 PR 27-SEP-2000; 2000US-235840P.  
 PR 27-SEP-2000; 2000US-235863P.  
 PR 28-SEP-2000; 2000US-236028P.  
 PR 28-SEP-2000; 2000US-236032P.  
 PR 28-SEP-2000; 2000US-236033P.  
 PR 28-SEP-2000; 2000US-236034P.  
 PR 28-SEP-2000; 2000US-236109P.  
 PR 28-SEP-2000; 2000US-236111P.  
 PR 29-SEP-2000; 2000US-236842P.  
 PR 29-SEP-2000; 2000US-236891P.  
 PR 02-OCT-2000; 2000US-237172P.  
 PR 02-OCT-2000; 2000US-237173P.  
 PR 02-OCT-2000; 2000US-237278P.  
 PR 02-OCT-2000; 2000US-237294P.  
 PR 02-OCT-2000; 2000US-237295P.  
 PR 03-OCT-2000; 2000US-237316P.  
 PR 03-OCT-2000; 2000US-237425P.  
 PR 03-OCT-2000; 2000US-237598P.  
 PR 03-OCT-2000; 2000US-237604P.  
 PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.

XX (AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;

DR WPI; 2002-198264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set

PS Claim 1; SEQ ID 2767; 44pp; English.

XX The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.

XX Sequence 2235 BP; 524 A; 632 C; 693 G; 386 T; 0 other;

XX Query Match 80.0%; Score 16.8; DB 24; Length 2235;  
 XX Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
 XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21

DB 1390 GTTCCCACTTGGGCCCATG 1371

RESULT 8

AAAC74249/c

ID AAC74249 standard; cDNA; 2304 BP.

XX

AC AAC74249;  
 XX 02-FEB-2001 (first entry)  
 DT  
 XX Human secreted protein gene 27 SEQ ID NO:37.

XX Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 KW neutropic; neuroprotective; antibacterial; virucide; fungicide; neoplasm;  
 KW ophthalmological; autoimmune disease; rheumatoid arthritis; angiogenesis;  
 KW hyperproliferative disorder; cardiovascular disorder; infection;  
 KW cerebrovascular disorder; nervous system disorder; ocular disorder;  
 KW wound healing; chemotaxis; ss.

XX Homo sapiens.

XX WO2000056754-A1.

XX 28-SEP-2000.

XX 16-MAR-2000; 2000WO-US06792.

XX 19-MAR-1999; 99US-0125362.

XX 10-DEC-1999; 99US-0169980.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen GA, Ruben SM, Komatsoulis G;

XX WPI; 2000-579483/54.

XX P-PSDB; AAB39205.

XX Isolated nucleic acid molecule encoding a human secreted protein is  
 PT used in preventing, treating or ameliorating a medical condition

XX Claim 1; Page 354-355; 434pp; English.

XX The polynucleotide sequences given in AAC74223-C74279 encode the human  
 CC secreted proteins represented in AAB39179-B39226. Sequences  
 CC AAB39227-B39308 are alternative proteins encoded by the genes, and also  
 CC protein sequences with which they share homology. The proteins have  
 CC activities based on the tissues and cells in which they are expressed.  
 CC Examples of activities include: immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;  
 CC cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; and ophthalmological. The human secreted proteins,  
 CC polynucleotides, antagonists and agonists of the invention may be useful  
 CC in the treatment, prevention, and/or diagnosis of various disease,  
 CC disorders and conditions such as autoimmune diseases e.g. rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The polypeptides can  
 CC also be used to aid wound healing and epithelial cell proliferation, to  
 CC regenerate tissues, maintain organs before transplantation, in  
 CC chemotaxis and as a food additive or preservative e.g. to increase  
 CC storage capabilities. Sequences AAC74214-C74222 and AAB39178 are used  
 CC during the isolation and characterisation of the genes of the invention.

XX Sequence 2304 BP; 692 A; 388 C; 472 G; 746 T; 6 other;

XX Query Match 80.0%; Score 16.8; DB 21; Length 2304;  
 XX Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
 XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTTCACACTTGGGCCCAT 20

DB 1894 AGTTCACACTTGGGCCCAT 1875

RESULT 9

AA158283/c

ID AAI58283 standard; cDNA; 2487 BP.  
AC AAI58283;  
XX 22-OCT-2001 (first entry)  
DT  
XX Human polynucleotide SEQ ID NO 486.  
DE  
XX Human: neurotrophic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX WO200153312-A1.  
PN 26-JUL-2001.  
XX 26-DEC-2000; 2000WO-US34253.  
PF  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-052317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0820312.  
PR 03-AUG-2000; 2000US-0853450.  
PR 14-SEP-2000; 2000US-062191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAM39127.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
PT  
XX Claim 1; SEQ ID NO 486; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM4213) with neurotrophic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2487 BP; 619 A; 678 C; 759 G; 431 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 2487;  
Best Local Similarity 90.0%; Pred. NO. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTGGGGCCCATG 21  
||||| ||||||| |||  
Db 1624 GTTCCCACTGGGGCCCATG 1605

RESULT 10  
ABK43576/C  
ID ABK43576 standard; cDNA; 2745 BP.  
XX  
XX AC ABK43576;  
XX  
XX 05-JUN-2002 (first entry)  
DT  
XX DNA encoding novel central nervous system protein #156.  
DE  
XX Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy; gene; ss.  
XX  
OS Homo sapiens.  
XX WO200155318-A2.  
PN  
XX 02-AUG-2001.  
PD  
XX 17-JAN-2001; 2001WO-US01332.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0228287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.



DT 22-OCT-2001 (first entry)  
 XX Human polynucleotide SEQ ID NO 4058.  
 XX  
 XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200153312-A1.  
 PN  
 XX 26-JUL-2001.  
 PD  
 XX 26-DEC-2000; 2000WO-US34263.  
 XX  
 XX 21-JAN-2000; 2000US-0488725.  
 PR  
 XX 25-APR-2000; 2000US-0552317.  
 PR  
 XX 09-JUL-2000; 2000US-0598042.  
 PR  
 XX 19-JUL-2000; 2000US-0620312.  
 PR  
 XX 03-AUG-2000; 2000US-0653450.  
 PR  
 XX 14-SEP-2000; 2000US-0662191.  
 PR  
 XX 19-OCT-2000; 2000US-0693036.  
 PR  
 XX 29-NOV-2000; 2000US-0727344.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 XX WPI: 2001-442253/47.  
 DR  
 XX P-PSDB; AAM40913.  
 DR  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PT  
 XX Claim 1; SEQ ID NO 4058; 10078pp; English.  
 PS  
 XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 XX Sequence 3443 BP; 921 A; 846 C; 940 G; 736 T; 0 other;  
 SQ  
 Query Match 80.0%; Score 16.8; DB 22; Length 3443;  
 Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 GTTCCCACTTGGGCCCCATG 21  
 DB 1627 GTTCCCACTTGGGCCCCATG 1608  
 RESULT 12  
 AAF54792/C  
 ID AAF54792 standard; cDNA: 5762 BP.

XX AAF54792;  
 AC  
 XX 15-MAY-2001 (first entry)  
 DT  
 XX Nucleotide sequence of a murine ABCA transporter.  
 DE  
 XX ABCA transporter; high-density lipoprotein cholesterol;  
 KW ABCA transporter; ss.  
 KW ABCA transporter; ss.  
 XX  
 OS Mus musculus.  
 XX  
 XX WO200109314-A1.  
 PN  
 XX 08-FEB-2001.  
 PD  
 XX 28-JAN-2000; 2000WO-FR00209.  
 XX  
 XX 30-JUL-1999; 99FR-0009926.  
 PR  
 XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PA  
 XX Chimini G;  
 PI  
 XX WPI: 2001-182953/18.  
 DR  
 XX Selecting agents that modulate ABCA transporters, useful e.g. for  
 PT normalizing serum cholesterol levels, comprises using transgenic  
 PT animals with an inactive ABCA gene allele  
 PS Claim 12; Page 78-80; 113pp; French.  
 PS  
 XX The specification describes a method for selecting or screening agents  
 CC that modulate ABCA transporters. The method comprises using non-human  
 CC recombinant mammals with an inactivated allele of the gene encoding the  
 CC ABCA transporter, or cells with an inactivated allele of ABCA  
 CC transporter, from any tissue of the recombinant mammal, preferably with  
 CC an allele truncated in one or both exons corresponding to the first  
 CC and/or second ATP-binding cassettes (NB1 or 2). Compounds that stimulate  
 CC ABCA transporters may be useful for increasing (normalizing) serum  
 CC levels of high-density lipoprotein cholesterol. The present sequence  
 CC represents the cDNA sequence of a murine ABCA transporter.  
 XX  
 XX Sequence 5762 BP; 1135 A; 1690 C; 1692 G; 1245 T; 0 other;  
 SQ  
 Query Match 80.0%; Score 16.8; DB 22; Length 5762;  
 Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 GTTCCCACTTGGGCCCCATG 21  
 DB 2038 GTTCCCACTTGGGCCCCATG 2019  
 RESULT 13  
 AAF54812/C  
 ID AAF54812 standard; cDNA: 6607 BP.  
 XX  
 AC AAF54812;  
 XX  
 XX 15-MAY-2001 (first entry)  
 DT  
 XX Nucleotide sequence of a murine ABCA (ABCA7) transporter.  
 DE  
 XX ABCA transporter; high-density lipoprotein cholesterol;  
 KW ABCA transporter; ABCA7 transporter; ss.  
 KW  
 XX Mus musculus.  
 OS  
 XX WO200109314-A1.  
 PN  
 XX 08-FEB-2001.  
 PD

XX PF 28-JAN-2000; 2000WO-FR00209.  
 XX PR 30-JUL-1999; 99FR-0009926.  
 XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Chimini G;  
 XX DR WPI; 2001-182953/18.  
 XX PT Selecting agents that modulate ABCA transporters, useful e.g. for  
 PT normalizing serum cholesterol levels, comprises using transgenic  
 PT animals with an inactive ABCA gene allele  
 XX PS Claim 12; Fig 22B; 113pp; French.  
 XX CC The specification describes a method for selecting or screening agents  
 CC that modulate ABCA transporters. The method comprises using non-human  
 CC recombinant mammals with an inactivated allele of the gene encoding the  
 CC ABCA transporter, or cells with an inactivated allele of ABCA  
 CC transporter, from any tissue of the recombinant mammal, preferably with  
 CC an allele truncated in one or both exons corresponding to the first  
 CC and/or second ATP-binding cassettes (NBD1 or 2). Compounds that stimulate  
 CC ABCA transporters may be useful for increasing (normalizing) serum  
 CC levels of high-density lipoprotein cholesterol. The present sequence  
 CC represents the cDNA sequence of a murine ABCAX (ABCA7) transporter.  
 XX SQ Sequence 6607 BP; 1307 A; 1954 C; 1924 G; 1422 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 6607;  
 Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCACACTGGGGCCCATG 21  
 ||||| || |||||  
 Db 2755 GTTCCGCTCGGGGCCCATG 2736

RESULT 14  
 AAF54793/c  
 ID AAF54793 standard; DNA; 14044 BP.  
 XX AC AAF54793;  
 XX DT 15-MAY-2001 (first entry)  
 XX DE Nucleotide sequence of a murine ABCAX transporter.  
 XX KW ABCA transporter; high-density lipoprotein cholesterol;  
 XX KW ABCAX transporter; ss.  
 XX OS Mus musculus.  
 XX PN WO200109314-A1.  
 XX PD 08-FEB-2001.  
 XX PF 28-JAN-2000; 2000WO-FR00209.  
 XX PR 30-JUL-1999; 99FR-0009926.  
 XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Chimini G;  
 XX DR WPI; 2001-182953/18.  
 XX PT Selecting agents that modulate ABCA transporters, useful e.g. for  
 PT normalizing serum cholesterol levels, comprises using transgenic  
 PT animals with an inactive ABCA gene allele

XX PS Claim 12; Page 80-84; 113pp; French.  
 XX CC The specification describes a method for selecting or screening agents  
 CC that modulate ABCA transporters. The method comprises using non-human  
 CC recombinant mammals with an inactivated allele of the gene encoding the  
 CC ABCA transporter, or cells with an inactivated allele of ABCA  
 CC transporter, from any tissue of the recombinant mammal, preferably with  
 CC an allele truncated in one or both exons corresponding to the first  
 CC and/or second ATP-binding cassettes (NBD1 or 2). Compounds that stimulate  
 CC ABCA transporters may be useful for increasing (normalizing) serum  
 CC levels of high-density lipoprotein cholesterol. The present sequence  
 CC represents the genomic sequence of a murine ABCAX transporter.  
 XX SQ Sequence 14044 BP; 2937 A; 3811 C; 4006 G; 3230 T; 60 other;

Query Match 80.0%; Score 16.8; DB 22; Length 14044;  
 Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCACACTGGGGCCCATG 21  
 ||||| || |||||  
 Db 5068 GTTCCGCTCGGGGCCCATG 5049

RESULT 15  
 AAF25499/c  
 ID AAF25499 standard; DNA; 23024 BP.  
 XX AC AAF25499;  
 XX DT 15-MAY-2001 (first entry)  
 XX DE Nucleotide sequence of a murine ABCAX (ABCA7) transporter.  
 XX KW ABCA transporter; high-density lipoprotein cholesterol;  
 XX KW ABCAX transporter; ss.  
 XX OS Mus musculus.  
 XX PN WO200109314-A1.  
 XX PD 08-FEB-2001.  
 XX PF 28-JAN-2000; 2000WO-FR00209.  
 XX PR 30-JUL-1999; 99FR-0009926.  
 XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Chimini G;  
 XX DR WPI; 2001-182953/18.  
 XX PT Selecting agents that modulate ABCA transporters, useful e.g. for  
 PT normalizing serum cholesterol levels, comprises using transgenic  
 PT animals with an inactive ABCA gene allele  
 XX PS Claim 12; Fig 22A; 113pp; French.  
 XX CC The specification describes a method for selecting or screening agents  
 CC that modulate ABCA transporters. The method comprises using non-human  
 CC recombinant mammals with an inactivated allele of the gene encoding the  
 CC ABCA transporter, or cells with an inactivated allele of ABCA  
 CC transporter, from any tissue of the recombinant mammal, preferably with  
 CC an allele truncated in one or both exons corresponding to the first  
 CC and/or second ATP-binding cassettes (NBD1 or 2). Compounds that stimulate  
 CC ABCA transporters may be useful for increasing (normalizing) serum  
 CC levels of high-density lipoprotein cholesterol. The present sequence  
 CC represents the genomic sequence of a murine ABCAX (ABCA7) transporter.  
 XX SQ Sequence 23024 BP; 5195 A; 6205 C; 6348 G; 5266 T; 10 other;

Query Match 80.0%; Score 16.8; DB 22; Length 23024;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GTTCCCACTGGGGCCCATG 21  
|||||.||| |||||  
Db 12681 GTTCCCGCTCGGGCCCATG 12662

Search completed: July 8, 2003, 02:18:50  
Job time : 133.941 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 28.9902 Seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-6

Perfect score: 21

Sequence: 1 agttccacttggggcccatg 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, NA: \*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	75.2	437	1	US-08-631-200-3
2	15.8	75.2	437	1	US-08-631-200-4
3	15.8	75.2	437	1	US-08-631-200-5
4	15.8	75.2	437	1	US-08-829-553-3
5	15.8	75.2	437	1	US-08-829-553-4
6	15.8	75.2	437	1	US-08-829-553-5
7	15.8	75.2	437	2	US-08-922-267A-3
8	15.8	75.2	437	2	US-08-922-267A-4
9	15.8	75.2	437	2	US-08-922-267A-5
10	15.8	75.2	437	2	US-08-936-707A-3
11	15.8	75.2	437	2	US-08-936-707A-4
12	15.8	75.2	437	2	US-08-936-707A-5
13	15.8	75.2	437	2	US-08-936-707A-3
14	15.8	75.2	437	2	US-08-936-706A-4
15	15.8	75.2	437	2	US-08-936-706A-5
16	15.8	75.2	437	3	US-09-248-203-3
17	15.8	75.2	437	3	US-09-248-203-4
18	15.8	75.2	437	3	US-09-248-203-5
19	15.8	75.2	437	4	US-09-406-071-3
20	15.8	75.2	437	4	US-09-406-071-4
21	15.8	75.2	437	4	US-09-406-071-5
22	15.8	75.2	480	1	US-08-630-592-5
23	15.8	75.2	480	1	US-08-714-991-5
24	15.8	75.2	480	3	US-09-032-365A-5
25	15.8	75.2	699	4	US-09-328-111-190
26	15.8	75.2	12394	4	US-09-488-856A-10
27	15.4	73.3	578	4	US-09-385-982-118

28	15.4	73.3	647	4	US-09-328-111-243
29	15.2	72.4	427	4	US-09-397-787-137
30	15.2	72.4	1815	4	US-09-042-785A-24
31	15.2	72.4	1868	1	US-08-392-367B-1
32	15.2	72.4	1868	3	US-08-893-467A-1
33	15.2	72.4	2186	3	US-08-959-382-1
34	15.2	72.4	2419	2	US-08-765-662-13
35	15.2	72.4	2419	5	PCT-US95-08745-13
36	15.2	72.4	2612	4	US-09-042-785A-3
37	15.2	72.4	2638	4	US-09-042-785A-22
38	15.2	72.4	3474	4	US-09-527-236A-1
39	15.2	72.4	6405	4	US-09-281-481A-18
40	14.8	70.5	272	1	US-08-362-670B-14
41	14.8	70.5	272	3	US-08-333-576C-14
42	14.8	70.5	272	4	US-08-808-324-14
43	14.8	70.5	272	5	PCT-US94-14030A-14
44	14.8	70.5	382	3	US-09-035-648-10
45	14.8	70.5	382	4	US-09-001-951-10

## ALIGNMENTS

RESULT 1  
US-08-631-200-3  
; Sequence 3, Application US/08631200  
; Patent No. 5646040  
; GENERAL INFORMATION:  
; APPLICANT: Kleyon, Patrick W.  
; APPLICANT: Moore, Karen J.  
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
; REFERENCE TO PARENT APPLICATION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDING ADDRESSES:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/631,200  
; FILING DATE: 12-APR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-057  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 437 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-631-200-3

Query Match 75.2%; Score 15.8; DB 1; Length 437;  
Best Local Similarity 89.5%; Pred No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| |||||  
Db 343 TTCCCTCTTGGGCCCATG 361

RESULT 2  
 US-08-631-200-4  
 ; Sequence 4, Application US/08631200  
 ; Patent No. 5646040  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Klynn, Patrick W.  
 ; APPLICANT: Moore, Karen J.  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
 ; NUMBER OF SEQUENCES: 59  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/631/200  
 ; FILING DATE: 12-APR-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7853-057  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 437 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: DNA (genomic)  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-631-200-4

Query Match 75.2%; Score 15.8; DB 1; Length 437;  
 Best Local Similarity 89.5%; Pred. No. 50;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCCCACTTGGGCCCATG 21  
 Db 343 TTCCCTCTGTGGCCCATG 361

RESULT 3  
 US-08-631-200-5  
 ; Sequence 5, Application US/08631200  
 ; Patent No. 5646040  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Klynn, Patrick W.  
 ; APPLICANT: Moore, Karen J.  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
 ; NUMBER OF SEQUENCES: 59  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/631/200  
 ; FILING DATE: 12-APR-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7853-057  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 437 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: cDNA  
 ; US-08-631-200-5

Query Match 75.2%; Score 15.8; DB 1; Length 437;  
 Best Local Similarity 89.5%; Pred. No. 50;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCCCACTTGGGCCCATG 21  
 Db 343 TTCCCTCTGTGGCCCATG 361

RESULT 4  
 US-08-829-553-3  
 ; Sequence 3, Application US/08829553  
 ; Patent No. 5817762  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Klynn, Patrick W.  
 ; APPLICANT: Moore, Karen J.  
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
 ; NUMBER OF SEQUENCES: 59  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/829/553  
 ; FILING DATE: 28-MAR-1997  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/631,200  
 ; FILING DATE: 12-APR-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7853-057  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 437 base pairs

;; TYPE: nucleic acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: DNA (genomic)  
US-08-829-553-3

Query Match 75.2%; Score 15.8; DB 1; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTGTGGCCCATG 361

RESULT 5  
US-08-829-553-4  
; Sequence 4, Application US/08829553  
; Patent No. 5817762  
; GENERAL INFORMATION:  
; APPLICANT: Kley, Patrick W.  
; APPLICANT: Moore, Karen J.  
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
; NUMBER OF SEQUENCES: 59  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/829,553  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/631,200  
; FILING DATE: 12-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-057  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 437 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-829-553-4

Query Match 75.2%; Score 15.8; DB 1; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTGTGGCCCATG 361

RESULT 6  
US-08-829-553-5  
; Sequence 5, Application US/08829553

;; Patent No. 5817762  
;; GENERAL INFORMATION:  
;; APPLICANT: Kley, Patrick W.  
;; APPLICANT: Moore, Karen J.  
;; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
;; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
;; NUMBER OF SEQUENCES: 59  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Pennie & Edmonds  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: U.S.A.  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/829,553  
;; FILING DATE: 28-MAR-1997  
;; CLASSIFICATION: 530  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/631,200  
;; FILING DATE: 12-APR-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Coruzzi, Laura A.  
;; REGISTRATION NUMBER: 30,742  
;; REFERENCE/DOCKET NUMBER: 7853-057  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 437 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: CDNA  
US-08-829-553-5

Query Match 75.2%; Score 15.8; DB 1; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTGTGGCCCATG 361

RESULT 7  
US-08-922-267A-3  
; Sequence 3, Application US/08922267A  
; Patent No. 5861239  
; GENERAL INFORMATION:  
; APPLICANT: Kley, Patrick W.  
; APPLICANT: Moore, Karen J.  
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,267A  
FILING DATE: 2-SEP-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/829,553  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/631,200  
FILING DATE: 12-APR-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-922-267A-3

Query Match 75.28; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTTGTGGCCCATG 361

RESULT 8  
US-08-922-267A-4  
Sequence 4, Application US/08922267A  
Patent No. 5861239  
GENERAL INFORMATION:  
APPLICANT: Kleyn, Patrick W.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,267A  
FILING DATE: 2-SEP-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/829,553  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-922-267A-4  
Query Match 75.28; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTTGTGGCCCATG 361

RESULT 9  
US-08-922-267A-5  
Sequence 5, Application US/08922267A  
Patent No. 5861239  
GENERAL INFORMATION:  
APPLICANT: Kleyn, Patrick W.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/922,267A  
FILING DATE: 2-SEP-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/829,553  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/631,200  
FILING DATE: 12-APR-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA  
US-08-922-267A-5

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTGTGGCCCATG 361

## RESULT 10

US-08-936-707A-3  
; Sequence 3, Application US/08936707A  
; Patent No. 5871931

## GENERAL INFORMATION:

APPLICANT: Kleynd, Patrick W.

APPLICANT: Moore, Karen J.

TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND

TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936,707A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-100

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 437 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-936-707A-3

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTGTGGCCCATG 361

## RESULT 11

US-08-936-707A-4

; Sequence 4, Application US/08936707A

; Patent No. 5871931

GENERAL INFORMATION:

APPLICANT: Kleynd, Patrick W.

APPLICANT: Moore, Karen J.

TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND

TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936,707A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-100

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 437 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-936-707A-4

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTGTGGCCCATG 361

## RESULT 12

US-08-936-707A-5  
; Sequence 5, Application US/08936707A  
; Patent No. 5871931

GENERAL INFORMATION:

APPLICANT: Kleynd, Patrick W.

APPLICANT: Moore, Karen J.

TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND

TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936,707A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7853-100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-936-707A-5

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTTGTGGCCCATG 361

RESULT 13  
US-08-936-706A-3  
Sequence 3, Application US/08936706A  
Patent No. 5876919  
GENERAL INFORMATION:  
APPLICANT: Kleyo, Patrick W.  
APPLICANT: Moore, Karen J.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,706A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-099  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-936-706A-3

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTTGTGGCCCATG 361

RESULT 14  
US-08-936-706A-4  
Sequence 4, Application US/08936706A  
Patent No. 5876919  
GENERAL INFORMATION:  
APPLICANT: Kleyo, Patrick W.  
APPLICANT: Moore, Karen J.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,706A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-099  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-936-706A-4

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTTGTGGCCCATG 361

RESULT 15  
US-08-936-706A-5  
Sequence 5, Application US/08936706A  
Patent No. 5876919  
GENERAL INFORMATION:  
APPLICANT: Kleyo, Patrick W.  
APPLICANT: Moore, Karen J.  
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF BODY WEIGHT DISORDERS, INCLUDING OBESITY  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCCCACTTGGGCCCATG 21  
||||| ||||| ||||| |||||  
Db 343 TTCCCTCTTGTGGCCCATG 361

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,706A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-099  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-936-706A-5

Query Match 75.2%; Score 15.8; DB 2; Length 437;  
Best Local Similarity 89.5%; Pred. No. 50;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 TTCCACTTGGGCCCATG 21  
|||||  
Db 343 TTCCCTCTTGGGCCCATG 361

Search completed: July 8, 2003, 09:31:00  
Job time : 30.0402 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 121.288 Seconds  
(without alignments)  
273.390 Million cell updates/sec.

Title: US-09-647-780A-6

Perfect score: 21

Sequence: 1 agttccacttgggcccatg 21

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.8	84.8	475	9	US-09-989-339-7
C 2	17.8	84.8	2639	9	US-09-989-339-1
C 3	17.4	82.9	493	9	US-09-918-995-31713
C 4	16.8	80.0	2235	9	US-10-171-311-39
C 5	16.8	80.0	2235	10	US-09-962-436-308
C 6	16.8	80.0	2487	9	US-10-037-270-160
C 7	16.8	80.0	3690	9	US-10-084-817-52
C 8	16.8	80.0	3690	12	US-10-044-090-448
C 9	16.8	80.0	6633	10	US-09-995-542-1
C 10	16.4	78.1	271	10	US-09-878-574-15243
C 11	16.4	78.1	500	9	US-09-918-995-30483
C 12	16.4	78.1	36221	9	US-09-954-556-29
C 13	16.2	77.1	402	10	US-09-864-761-14177
C 14	16.2	77.1	485	9	US-10-198-846-10268
C 15	16.2	77.1	1298	9	US-10-023-896-19
C 16	16.2	77.1	2127	9	US-10-004-551-17
C 17	16.2	77.1	2127	9	US-10-004-551-19
C 18	16.2	77.1	2143	9	US-10-004-551-23
C 19	16.2	77.1	15849	10	US-09-880-107-2362

C 20	16	76.2	211	10	US-09-783-590-145
C 21	15.8	75.2	437	10	US-09-814-986-3
C 22	15.8	75.2	437	10	US-09-814-986-4
C 23	15.8	75.2	437	10	US-09-814-986-5
C 24	15.8	75.2	699	10	US-09-879-536-190
C 25	15.8	75.2	9249	9	US-10-270-333-67
C 26	15.4	73.3	402	10	US-09-864-761-3270
C 27	15.4	73.3	486	9	US-09-918-995-9442
C 28	15.4	73.3	578	9	US-09-871-161-118
C 29	15.4	73.3	647	10	US-10-156-761-6794
C 30	15.4	73.3	2469	9	US-10-156-761-1
C 31	15.4	73.3	98829	9	US-09-864-761-18646
C 32	15.4	73.3	9025608	9	US-10-040-733-75
C 33	15.2	72.4	259	10	US-09-864-761-20166
C 34	15.2	72.4	286	9	US-09-918-995-37628
C 35	15.2	72.4	380	9	US-09-803-719-908
C 36	15.2	72.4	427	10	US-09-876-889-137
C 37	15.2	72.4	448	9	US-09-764-891-1580
C 38	15.2	72.4	462	9	US-09-918-995-28561
C 39	15.2	72.4	462	9	US-10-198-846-104
C 40	15.2	72.4	467	10	US-09-864-761-2463
C 41	15.2	72.4	538	9	US-09-918-995-27083
C 42	15.2	72.4	665	12	US-10-044-090-466
C 43	15.2	72.4	708	10	US-09-974-300-4450
C 44	15.2	72.4			
C 45	15.2	72.4			

## ALIGNMENTS

RESULT 1  
US-09-989-339-7/c  
Sequence 7, Application US/09989339  
Publication No. US2003008886A1  
GENERAL INFORMATION:  
APPLICANT: Falco, Saverio Carl  
APPLICANT: Famodu, Layo  
APPLICANT: Rafalski, Jan A.  
APPLICANT: Ramaker, Michael  
APPLICANT: Tarczynski, Mitchell C.  
APPLICANT: Thorpe, Catherine  
TITLE OF INVENTION: PLANT METHIONINE SYNTHASE GENE AND METHODS FOR INCREASING THE  
FILE REFERENCE: BB-1067-B  
CURRENT APPLICATION NUMBER: 08/703,829  
PRIOR FILING DATE: 1996-08-27  
NUMBER OF SEQ ID NOS: 55  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 7  
LENGTH: 475  
TYPE: DNA  
ORGANISM: Triticum aestivum  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (344)  
OTHER INFORMATION: n = A, C, G, or T  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (367)  
OTHER INFORMATION: n = A, C, G, or T  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (433)  
OTHER INFORMATION: n = A, C, G, or T  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (452)  
OTHER INFORMATION: n = A, C, G, or T  
FEATURE:  
NAME/KEY: unsure

Sequence 145, App  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 5, Appli  
Sequence 190, App  
Sequence 67, Appl  
Sequence 3270, Ap  
Sequence 9442, Ap  
Sequence 118, App  
Sequence 243, App  
Sequence 6794, Ap  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 18646, A  
Sequence 20166, A  
Sequence 75, Appl  
Sequence 37628, A  
Sequence 908, App  
Sequence 137, App  
Sequence 1580, App  
Sequence 28561, A  
Sequence 104, App  
Sequence 27083, A  
Sequence 466, App  
Sequence 4450, App

us-09-647-780a-6.rnpb

Wed Jul 9 09:33:39 2003

Query Match 82.9%; Score 17.4; DB 9; Length 493;  
Best Local Similarity 90.0%; Pred. No. 23;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGGCCCATG 21  
Db 433 GATCCCACTTGGGGCCCATG 452

RESULT 4

US-10-171-311-39/c  
; Sequence 39, Application US/10171311  
; Publication No. US20030087270A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Chen, Yan  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Monahan, John  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Glatt, Karen  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Hoerssh, Sebastian  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; TITLE OF INVENTION: OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035  
; CURRENT APPLICATION NUMBER: US/10/171,311  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 39  
; LENGTH: 2235  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-171-311-39

Query Match 80.0%; Score 16.8; DB 9; Length 2235;  
Best Local Similarity 90.0%; Pred. No. 40;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGGCCCATG 21  
Db 1390 GTTCCCACTTGGGGCCCATG 1371

RESULT 5

US-09-962-436-308/c  
; Sequence 308, Application US/09962436  
; Patent No. US20020081301A1  
; GENERAL INFORMATION:  
; APPLICANT: Soppet, Daniel  
; TITLE OF INVENTION: Cancer  
; TITLE OF INVENTION: Sets  
; FILE REFERENCE: 689290-75  
; CURRENT APPLICATION NUMBER: US/09/962,436  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,082  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/234,924  
; PRIOR FILING DATE: 2000-09-25  
; NUMBER OF SEQ ID NOS: 568  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 308  
; LENGTH: 2235  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-962-436-308

LOCATION: (473)...(474)  
OTHER INFORMATION: n = A, C, G, or T  
US-09-989-339-7

Query Match 84.8%; Score 17.8; DB 9; Length 475;  
Best Local Similarity 90.5%; Pred. No. 15;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTCCCACTTGGGGCCCATG 21  
Db 124 AGTCCCTCTTGGGGCCCATG 104

RESULT 2

US-09-989-339-1/c  
; Sequence 1, Application US/09989339  
; Publication No. US2003008886A1  
; GENERAL INFORMATION:  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Famodu, Layo  
; APPLICANT: Rafalski, Jan A.  
; APPLICANT: Ramaker, Michael  
; APPLICANT: Tarczynski, Mitchell C.  
; TITLE OF INVENTION: PLANT METHIONINE SYNTHASE GENE AND METHODS FOR INCREASING THE  
; TITLE OF INVENTION: METHIONINE CONTENT OF THE SEEDS OF PLANTS  
; FILE REFERENCE: BB-1067-B  
; CURRENT APPLICATION NUMBER: US/09/989,339  
; CURRENT FILING DATE: 2002-05-31  
; PRIOR APPLICATION NUMBER: 08/7703,829  
; PRIOR FILING DATE: 1996-08-27  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 1  
; LENGTH: 2639  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-989-339-1

Query Match 84.8%; Score 17.8; DB 9; Length 2639;  
Best Local Similarity 90.5%; Pred. No. 13;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTCCCACTTGGGGCCCATG 21  
Db 147 AGTCCCTCTTGGGGCCCATG 127

RESULT 3

US-09-918-995-31713  
; Sequence 31713, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 31713  
; LENGTH: 493  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(493)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31713

```
Query Match
Best Local Similarity 80.0%; Score 16.8; DB 10; Length 2235;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21
Db 1390 GTTCCCACTTGGGCCCATG 1371

RESULT 6
US-10-037-270-160/c
; Sequence 160, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoye T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; PRIOR FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: PL_FL_Genes Version 1.0
; SEQ ID NO 160
; LENGTH: 2487
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (235)..(2469)
; US-10-037-270-160

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 2487;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21
Db 1624 GTTCCCACTTGGGCCCATG 1605

RESULT 7
US-10-084-817-52/c
; Sequence 52, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Pion
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 12; Length 3690;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21
Db 1597 GTTCCCACTTGGGCCCATG 1578

RESULT 9
US-09-995-542-1/c
; Sequence 1, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
; APPLICANT: Shutter, John
; APPLICANT: Ulias, Laarni
; TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
; FILE REFERENCE: 00-658-A
; CURRENT APPLICATION NUMBER: US/09/995,542
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,520
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver: 2.0
; SEQ ID NO 1
; LENGTH: 6633
; TYPE: DNA
; ORGANISM: Mus musculus

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 3690;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21
Db 1597 GTTCCCACTTGGGCCCATG 1578

RESULT 8
US-10-044-090-448/c
; Sequence 448, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandhan
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 448
; LENGTH: 3690
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 2314132CB1
; US-10-044-090-448

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 3690;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21
Db 1597 GTTCCCACTTGGGCCCATG 1578

RESULT 8
US-10-044-090-448/c
; Sequence 448, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandhan
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 448
; LENGTH: 3690
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 2314132CB1
; US-10-044-090-448

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 3690;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTTGGGCCCATG 21
Db 1597 GTTCCCACTTGGGCCCATG 1578

RESULT 9
US-09-995-542-1/c
; Sequence 1, Application US/09995542
; Patent No. US20020127647A1
; GENERAL INFORMATION:
; APPLICANT: Shutter, John
; APPLICANT: Ulias, Laarni
; TITLE OF INVENTION: ATP-Binding Cassette Transporter-Like Molecules and
; FILE REFERENCE: 00-658-A
; CURRENT APPLICATION NUMBER: US/09/995,542
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 60/253,520
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver: 2.0
; SEQ ID NO 1
; LENGTH: 6633
; TYPE: DNA
; ORGANISM: Mus musculus
```

Wed Jul 9 09:33:39.2003

us-09-647-780a-6.rnpb

FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: (1)..(138)  
NAME/KEY: CDS  
LOCATION: (1)..(6504)  
US-09-995-542-1

Query Match 80.0%; Score 16.8; DB 10; Length 6633;  
Best Local Similarity 90.0%; Pred. No. 37;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTTCCCACTGGGGCCCATG 21  
||||||| || |||||  
Db 2755 GTTCCCGCTCGGGCCCATG 2736

RESULT 10  
US-09-878-574-15243  
; Sequence 15243, Application US/09878574  
; Patent No. US20020110548A1  
; GENERAL INFORMATION:  
; APPLICANT: Byrum, Joseph R.  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Thompson, Michael D.  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(15401)B  
; CURRENT APPLICATION NUMBER: US/09/878,574  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/333,535  
; PRIOR FILING DATE: 1999-06-14  
; NUMBER OF SEQ ID NOS: 15775  
; SEQ ID NO 15243  
; LENGTH: 271  
; TYPE: DNA  
; ORGANISM: Glycine max  
; OTHER INFORMATION: Clone ID: 701069847H1  
US-09-878-574-15243

Query Match 78.1%; Score 16.4; DB 10; Length 271;  
Best Local Similarity 94.4%; Pred. No. 74;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GTTCCCACTGGGGCCCA 19  
||||||| |||||  
Db 136 GTTCCCACTGGGGCACA 153

RESULT 11  
US-09-918-995-30483/c  
; Sequence 30483, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 30483  
; LENGTH: 500  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(500)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-30483

Query Match 78.1%; Score 16.4; DB 9; Length 500;  
Best Local Similarity 94.4%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGTTCACCTGGGGCCC 18  
||||||| |||||  
Db 103 AGTTCACCTGGGGCCC 86

RESULT 12  
US-09-954-556-29/c  
; Sequence 29, Application US/09954556  
; Publication No. US20030078219A1  
; GENERAL INFORMATION:  
; APPLICANT: Brett P. Monia  
; APPLICANT: Susan M. Freier  
; APPLICANT: Scott Cooper  
; TITLE OF INVENTION: ANTISENSE MODULATION OF FIBROBLAST GROWTH FACTOR RECEPTOR 2 EX  
; FILE REFERENCE: RTS-0250  
; CURRENT APPLICATION NUMBER: US/09/954,556  
; CURRENT FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 108  
; SEQ ID NO 29  
; LENGTH: 36221  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(36221)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-954-556-29

Query Match 78.1%; Score 16.4; DB 9; Length 36221;  
Best Local Similarity 94.4%; Pred. No. 52;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TCCCACTGGGGCCCATG 21  
||||||| |||||  
Db 14675 TCCCACTGGGGCCCATG 14658

RESULT 13  
US-09-864-761-14177/c  
; Sequence 14177, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 14177  
LENGTH: 402  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC010872.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
US-09-864-761-14177

Query Match 77.1%; Score 16.2; DB 10; Length 402;  
Best Local Similarity 85.7%; Pred. No. 90;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGTCCCACTGGGGCCCATG 21  
|| ||||| ||||| |||||  
DB 249 AGGCCCACTGGGGCCCATG 229

RESULT 14  
US-10-198-846-10268/c  
Sequence 10268, Application US/10198846  
Publication No. US2003009974A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Xu, Yongyao  
APPLICANT: Wang, Youzhen  
APPLICANT: Steinmann, Kathleen  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
TITLE OF INVENTION: THERAPY OF BREAST CANCER  
FILE REFERENCE: MRI-049  
CURRENT APPLICATION NUMBER: US/10/198,846  
CURRENT FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/306,220  
PRIOR FILING DATE: 2001-07-18  
NUMBER OF SEQ ID NOS: 14084  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10268  
LENGTH: 485  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 480, 481, 482, 483, 484, 485  
OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-10268

Query Match 77.1%; Score 16.2; DB 9; Length 485;  
Best Local Similarity 85.7%; Pred. No. 89;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGTCCCACTGGGGCCCATG 21

DB 227 AGTACCCATTGGGGCCCATG 207  
||| |||| ||||| ||||| |||||

RESULT 15  
US-10-023-896-19  
Sequence 19, Application US/10023896  
Publication No. US20030027776A1  
GENERAL INFORMATION:  
APPLICANT: Victor Roschke  
TITLE OF INVENTION: 29 Human Cancer Associated Proteins  
FILE REFERENCE: PA004P1  
CURRENT APPLICATION NUMBER: US/10/023,896  
CURRENT FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: unassigned  
PRIOR FILING DATE: 2001-12-21  
PRIOR APPLICATION NUMBER: PCT/US00/23794  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/152,296  
PRIOR FILING DATE: 1999-09-03  
PRIOR APPLICATION NUMBER: 60/158,003  
PRIOR FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 1298  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-023-896-19

Query Match 77.1%; Score 16.2; DB 9; Length 1298;  
Best Local Similarity 85.7%; Pred. No. 83;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGTCCCACTGGGGCCCATG 21  
||||| ||||| ||||| |||||  
DB 74 AGTCCCAAGTGGGGCCCATG 94

Search completed: July 9, 2003, 02:21:51  
Job time : 126.338 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1064.65 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-6  
Perfect score: 21  
Sequence: 1 agttccacttggggcccatg 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_man:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.4	92.4	600	13	BI989738 4044-75 M
C 2	18.4	87.6	204	12	BF656002 FMI_42_FI
C 3	18.4	87.6	545	13	BI90926 BI90926
C 4	18.4	85.7	407	14	R07454 ye96f09.s1
C 5	17.8	84.8	115	12	BF474621 WHE2108.B
C 6	17.8	84.8	136	13	BJ280551 BJ280551

C 7	17.8	84.8	139	13	BJ280541
C 8	17.8	84.8	146	10	AV939865
C 9	17.8	84.8	216	13	BJ235482
C 10	17.8	84.8	234	12	BG050044 FMI_65_G0
C 11	17.8	84.8	238	14	D41509
C 12	17.8	84.8	263	12	BG465041
C 13	17.8	84.8	265	14	D40937
C 14	17.8	84.8	270	9	AV167513
C 15	17.8	84.8	283	14	D47961
C 16	17.8	84.8	293	12	BF481816
C 17	17.8	84.8	304	10	BE638127 WHE0995-0
C 18	17.8	84.8	304	14	C73949
C 19	17.8	84.8	312	9	AL507110
C 20	17.8	84.8	317	13	BJ469066
C 21	17.8	84.8	333	12	BG050832 FMI_71_D0
C 22	17.8	84.8	337	10	BE500035 WHE09977_D
C 23	17.8	84.8	339	13	BJ225552
C 24	17.8	84.8	350	12	BF481363 FMI_18_B0
C 25	17.8	84.8	352	12	BF483281 WHE1791.A
C 26	17.8	84.8	353	14	C27148
C 27	17.8	84.8	359	12	BG050316 FMI_53_E0
C 28	17.8	84.8	361	13	BI098822
C 29	17.8	84.8	364	13	BJ280025
C 30	17.8	84.8	366	13	BI098823
C 31	17.8	84.8	367	13	BI163201
C 32	17.8	84.8	368	10	BE519036
C 33	17.8	84.8	368	12	BG050750
C 34	17.8	84.8	368	17	AZ902604
C 35	17.8	84.8	375	9	AJ433505
C 36	17.8	84.8	378	10	BE414865
C 37	17.8	84.8	382	12	BF481918
C 38	17.8	84.8	386	14	D40154
C 39	17.8	84.8	387	12	BF421154
C 40	17.8	84.8	388	13	BI185439
C 41	17.8	84.8	390	12	BG050135
C 42	17.8	84.8	392	13	BI183732
C 43	17.8	84.8	394	9	AU096450
C 44	17.8	84.8	395	12	BG049991
C 45	17.8	84.8	398	10	AV940980

## ALIGNMENTS

RESULT 1  
BI989738/c  
LOCUS BI989738  
DEFINITION 4044-75 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
600 bp mRNA linear EST 20-DEC-2001  
mRNA sequence.  
ACCESSION BI989738  
VERSION BI989738.1 GI:17960740  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 600)  
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329  
Location/Qualifiers  
1. 600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

FEATURES  
source  
Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329  
Location/Qualifiers  
1. 600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

information regarding entire library may be found at  
[http://pga.swned.edu/data/Libraries/microarray\\_cdna\\_libraries.htm](http://pga.swned.edu/data/Libraries/microarray_cdna_libraries.htm).

LOCUS	Bu190926	545 bp	mRNA	linear	EST 24-JAN-2002
DEFINITION	Bu190926	normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata Physcomitrella patens subsp. patens cDNA clone pphn1118 5', mRNA sequence.			
ACCESSION	Bu190926				
VERSION	Bu190926.1	GI:18358867			
KEYWORDS	EST				

SOURCE	ORGANISM
Physcomitrella patens subsp. patens.	
Physcomitrella patens subsp. patens	
Eukaryote; Viridiplantae; Streptophyta;	Embryophyta; Bryophyta;

AUTHORS	Fujita, T., Shin-i, T., Seki, M., Kamiya, A., Uchiyama, I., Nishiyama, T., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe M.
TITLE	Comparison of the moss <i>Physcomitrella patens</i> genome with flowering plants genome
JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information

National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp

A backbone of the vector is pBluescript II, that was in vivo excised from a modified LP8 phage vector (Mobi Tec Germany). XhoI digested-5' end of cDNA is ligated to SalI site of the vector.

the BamHI digested-3' end including poly-A tail is ligated to the vector, and the site of the vector. cDNA insert could be amplified with conventional T7 and T3 primers. This normalized full-length cDNA library was generated basically according to the method described in Genome Research 10, 1617-1630 (2000). Carninci, P. et al. Protonemata were blended by the POLYTRON, and then cultivated on the BCD medium containing 1mM NAA (naphthalene acetic acid) for 8 to 11 days under the continuous light.

FEATURES	source	Location/Qualifiers
		1. .545
		/organism="Physcomitrella patens subsp. patens"
		/db_xref="taxon:145481"
		/clone="pphnl118"
		/clone_lib="normalized full length cDNA library, chloronemata, caulonemata and rhizoid-like protonemata"

```

BASE COUNT      124 a   154 c   140 g   127 t
ORIGIN
Query Match      87.6%;   Score 18.4;   DB 13;   Length 545;
Best Local Similarity 95.0%;   Pred. No. 1.8e+02;
Matches 19;   Conservative. 0;   Mismatches 1;   Indels 0;   Gaps 0;

QY      1 AGTTCCTCCACTTGGGGCCCAT 20
          |||||  |||||  |||||
Db      172 AGTTCCTCTTGGGGCCCAT 153

RESULT 4
R07454
LOCUS

```

DEFINITION	Soares fetal liver spleen INFLS Homo sapiens cdna clone
ACCESSION	U05495
IMAGE	IMAGE:125609 3', mRNA sequence.
KEYWORDS	key words: liver, spleen, INFLS, Homo sapiens, cdna, clone
REFERENCES	1. Soares J, et al. (1995) J. Biol. Chem. 270:12560-12569
NOTES	
REMARKS	
FEATURES	1..407 bp. mRNA. linear. EST U05495
ORIGIN	no428
ORIGIN	ye98709.sl
ORIGIN	R07454

VERSION	R07454.1	GI:759377	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 407)		
AUTHORS	Hillier,L., Clark,N., Dubucque,T., Elliston,K., Hawkins,M., Holman M., Hultman,M., Kucaba,T., Le.M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston R., Williamson,A., Wohldmann,P. and Wilson,R. The WashU-Merck EST Project		
TITLE	Unpublished (1995)		
JOURNAL	Contact: Wilton RK		
COMMENT	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Insert Size: 1036 High quality sequence stops: 303 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 1036 Std Error: 0.00 Seq primer: -21ml3 High quality sequence stop: 303. Location/Qualifiers		
FEATURES	source		
	1..407	/organism="Homo sapiens" /db_xref="GDB:478154" /db_xref="taxon:9606" /clone="IMAGE:125609" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pTT73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' ACTCGAAGATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	104 a	85 c 72 g 142 t 4 others	
ORIGIN			
Query Match	85.7%	Score 18; DB 14; Length 407;	
Best Local Similarity	100.0%	Pred. No. 2.5e+02;	
Matches	18; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
Qy	3 TTCCCACTTGCGGCCCAT 20		
Db	290 TTCCCACTTGCGGCCCAT 307		
RESULT 5			
BF474621/c			
LOCUS	BF474621	115 bp mRNA linear EST 04-DEC-2000	
DEFINITION	WHE2108_B06_D12S Wheat salt-stressed crown cDNA library Triticum aestivum CDNA clone WHE2108_B06_D12, mRNA sequence.		
ACCESSION	BF474621		
VERSION	BF474621.1	GI:11543803	
KEYWORDS	EST.		
SOURCE	bread wheat.		
ORGANISM	Triticum aestivum		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum. 1 (bases 1 to 115)			
REFERENCE	Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J., Khanov,I., ...		

Lazo,G.R., Miller,R., Otto,C., Nguyen,H.T., Rausch,C.J., Seaton C.L., Simons,K., Tong,J.C. and Zhang,D. The structure and function of the expressed portion of the wheat genomes - Salt-stressed crown cDNA library Unpublished (2000) Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 5105595773 Fax: 5105595818 Email: andersn@pw.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stratagene SK primer. Location/Qualifiers			
FEATURES	source		
	1..115	/organism="Triticum aestivum" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="WHE2108_B06_D12" /clone_lib="Wheat salt-stressed crown cDNA library" /tissue_type="Crown" /dev_stage="Adult plant" /lab_host="E. coli SOLR" /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site.1: EcoRI; Site.2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J Dvorak Lab). Total RNA and poly(A) RNA were prepared from from crown tissue, equal portions of RNA were pooled from the two treatments, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Clone lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)." .	
BASE COUNT	22 a	35 c 34 g 24 t	
ORIGIN			
Query Match	84.8%	Score 17.8; DB 12; Length 115;	
Best Local Similarity	90.5%	Pred. No. 2e+02;	
Matches	19; Conservative	0; Mismatches, 2; Indels 0; Gaps 0;	
QY	1 AGTTCCTCCACTGGGCCCCATG 21		
Db	83 AGCTCCCTCTGTGGGCCCATG 63		
RESULT 6			
BJ280551/c			
LOCUS	BJ280551	136 bp mRNA linear EST 09-APR-2002	
DEFINITION	BJ280551 Y. Ogihara unpublished cDNA library, Wht_r Triticum aestivum CDNA clone whr8n23 5', mRNA sequence.		
ACCESSION	BJ280551		
VERSION	BJ280551.1	GI:20102609	
KEYWORDS	EST.		
SOURCE	bread wheat.		
ORGANISM	Triticum aestivum		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum. 1 (bases 1 to 136)			
REFERENCE	Ogihara,Y. and Murai,K. Expressed genes in Triticum aestivum Unpublished (2002) Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855		

Email: tshini@genes.nig.ac.jp.

# FEATURES

Location/Qualifiers  
1. 136  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="whr8n23"  
/clone\_lib="Y. Ogihara unpublished cDNA library, wh\_r"  
/tissue\_type="root"  
/dev\_stage="Feekes' scale 1"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J. Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Clonase lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

## BASE COUNT

31 a 28 c 49 g 28 t

## Query Match

Best Local Similarity 84.8%; Score 17.8; DB 13; Length 136;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTCCCTCTGGGCCCATG 21

Db 21 AGCTCCCTCTGGGCCCATG 1

## RESULT 7

### LOCUS

DEFINITION BJ280541 139 bp mRNA linear EST 09-APR-2002  
aestivum cDNA clone whr8n17 5', mRNA sequence.

### ACCESSION

VERSION BJ280541

### KEYWORDS

SOURCE EST.

### ORGANISM

Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

### REFERENCE

AUTHORS 1 (bases 1 to 139)

TITLE Ogihara, Y. and Mura, K.

JOURNAL Expressed genes in Triticum aestivum

COMMENT Unpublished (2002)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1. 139

/organism="Triticum aestivum"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

/clone="whr8n17"

/clone\_lib="Y. Ogihara unpublished cDNA library, wh\_r"

/tissue\_type="root"

/dev\_stage="Feekes' scale 1"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J. Dvorak Lab). Total RNA was prepared from sheath

tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Clonase lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

## BASE COUNT

31 a 37 c 46 g 25 t

## Query Match

Best Local Similarity 84.8%; Score 17.8; DB 13; Length 139;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTCCCTCTGGGCCCATG 21

Db 42 AGCTCCCTCTGGGCCCATG 22

## RESULT 8

### LOCUS

DEFINITION AV939865 146 bp mRNA linear EST 18-JAN-2002  
Hordeum vulgare subsp. spontaneum  
heading stage top three leaves Hordeum vulgare subsp. spontaneum  
cDNA clone bah24f19 5', mRNA sequence.

### ACCESSION

VERSION AV939865

### KEYWORDS

SOURCE EST.

### ORGANISM

Hordeum vulgare subsp. spontaneum.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae

### REFERENCE

AUTHORS 1 (bases 1 to 146)

TITLE Sato, K., Saisho, D. and Takeda, K.

JOURNAL Barley EST sequencing project in NIG and Okayama Univ

COMMENT Unpublished (2002)

Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

Location/Qualifiers

1. 146

/organism="Hordeum vulgare subsp. spontaneum"

/strain="H602"

/db\_xref="taxon:77009"

/clone="bah24f19"

/clone\_lib="K. Sato unpublished cDNA library, strain H602"

/tissue\_type="top three leaves"

/dev\_stage="adult, heading stage"

20 a 63 c 38 g 25 t

## BASE COUNT

20 a 63 c 38 g 25 t

## Query Match

Best Local Similarity 84.8%; Score 17.8; DB 10; Length 146;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGTCCCTCTGGGCCCATG 21

Db 122 AGCTCCCTCTGGGCCCATG 102

## RESULT 9

### LOCUS

DEFINITION BJ235482 216 bp mRNA linear EST 05-APR-2002  
aestivum cDNA clone whel6n16 5', mRNA sequence.

### ACCESSION

VERSION BJ235482



SULT 13  
0937/C  
CUS  
D40937      265 bp      mRNA      linear      EST 03-APR-2002  
RICS3131A Rice shoot Oryza sativa (japonica cultivar-group) cDNA,  
mRNA sequence.  
D40937  
CESSION D40937  
RSION D40937.1 GI:570088  
YWORDS  
EST.  
Oryza sativa (japonica cultivar-group)  
URCE  
ORGANISM Oryza sativa (japonica cultivar-group)

COMMENT

Contact: Chie Owa  
Genome Science Laboratory  
RIKEN  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145  
Fax: 81-298-36-9098  
Email: genome-res@rctc.riken.go.jp

Thermotransformation and thermocatalysis of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers

```

FEATURES              - RGR
Location/Qualifiers
1..283
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"
56 a      89 c    73 g    50 t    5 others
          Query Match      84.8%; Score 17.8; DB 14; Length 283;
          Best Local Similarity 90.5%; Pred No. 2.8e+02;
          Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AGTTCGCCACTGGGCCCCCATG 21
   ||| |||| |||||| ||||| ||| ..

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 220.098 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-7  
Perfect score: 20  
Sequence: 1 gctggaggattctctgtcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues.

Total number of hits satisfying chosen parameters: 4109280

Minimum DB-seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.:

1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pi.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	20	100.0	20	6	AX014707 Sequence
2	20	100.0	2765	6	AX014701 Sequence
3	20	100.0	174953	2	AC094732 Rattus no
4	19	95.0	157126	2	AC121182 Rattus no
5	19	95.0	179992	2	AC095165 Rattus no
6	18.4	92.0	1592	10	MUSHSB3
7	18.4	92.0	77333	2	AC022586 Homo sapi
8	18.4	92.0	163880	9	AC009505 Homo sapi
9	18.4	92.0	214433	2	AL627184 Mus muscu
10	18	90.0	110491	2	AC095501 Rattus no
11	18	90.0	158869	2	AC095500 Rattus no
12	17.4	87.0	2773	3	AY058691 Drosophil
13	17.4	87.0	29937	2	AC095847 Rattus no
14	17.4	87.0	54186	2	AC019884 Drosophil
15	17.4	87.0	67833	3	AC004564 Drosophil
16	17.4	87.0	167986	2	AC120994 Rattus no
17	17.4	87.0	174028	2	AC099354 Rattus no
18	17.4	87.0	179139	3	AC099307 Drosophil
19	17.4	87.0	313634	3	AE003454 Drosophil
20	17	85.0	566	6	AX164254 Sequence
21	17	85.0	639	9	F330116S11
22	17	85.0	6599	9	HS251507 Homo sapi
23	17	85.0	6822	6	AX401928 Sequence
24	17	85.0	6822	10	RNSCIII
25	17	85.0	8976	9	AF225986 Homo sapi
26	17	85.0	9112	6	AX164235 Sequence
27	17	85.0	9112	6	AX164236 Sequence
28	17	85.0	9123	9	AF225987 Homo sapi
29	17	85.0	129241	9	AL353660 Human DNA
30	17	85.0	151018	9	AC073636 Homo sapi
31	17	85.0	168493	9	AC013463 Homo sapi
32	17	85.0	172135	2	AC074009 Homo sapi
33	17	85.0	182448	2	AC124034 Rattus no
34	17	85.0	183561	2	AC012452 Homo sapi
35	17	85.0	220831	2	AC105645 Rattus no
36	16.8	84.0	1627	10	BC012715 Mus muscu
37	16.8	84.0	1661	9	BC007917 Homo sapi
38	16.8	84.0	1769	10	MADES5
39	16.8	84.0	2285	10	RNDES
40	16.8	84.0	3420	11	G54303
41	16.8	84.0	3420	11	G54304
42	16.8	84.0	3546	10	MUSENT01
43	16.8	84.0	3577	10	AF228347
44	16.8	84.0	7794	10	MMU250633
45	16.8	84.0	19391	10	MMDESMINP

# ALIGNMENTS

RESULT 1  
AX014707  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AX014707  
Sequence 7 from Patent WO9953077.  
AX014707  
AX014707.1 GI:10040980

synthetic construct.  
synthetic construct  
artificial sequences.  
1 (bases 1 to 20)  
Schwartz J.C., Gros, C., Oulmet, T., Rose, C., Bonhomme, M.C. and  
Facchinetti, P.  
Novel nep 11 membrane metalloprotease and its use for screening  
inhibitors useful in therapy

20 bp  
DNA  
linear  
PAT 07-SEP-2000



Center clone name: CH230-516  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329first call to  
findPhrapList  
Consensus quality: 152255 bases at least Q40  
Consensus quality: 158448 bases at least Q30  
Consensus quality: 16461 bases at least Q20  
Estimated insert size: 155965; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hsc.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hsc.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 56 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 13782: contig of 13782 bp in length  
\* 13783 13882: gap of unknown length  
\* 13883 23287: contig of 9405 bp in length  
\* 23288 23387: gap of unknown length  
\* 23388 28081: contig of 4694 bp in length  
\* 28082 28181: gap of unknown length  
\* 28182 33807: contig of 5626 bp in length  
\* 33808 33971: gap of unknown length  
\* 33972 39271: contig of 5364 bp in length  
\* 39272 39371: gap of unknown length  
\* 39372 44270: contig of 4899 bp in length  
\* 44271 44370: gap of unknown length  
\* 44371 47723: contig of 3353 bp in length  
\* 47724 47823: gap of unknown length  
\* 47824 53427: contig of 5604 bp in length  
\* 53428 53527: gap of unknown length  
\* 53528 57303: contig of 3776 bp in length  
\* 57304 57403: gap of unknown length  
\* 57404 62018: contig of 4615 bp in length  
\* 62019 62118: gap of unknown length  
\* 62119 66676: contig of 4558 bp in length  
\* 66677 66776: gap of unknown length  
\* 66777 70201: contig of 3425 bp in length  
\* 70202 70301: gap of unknown length  
\* 70302 74082: contig of 3781 bp in length  
\* 74083 74182: gap of unknown length  
\* 74183 78306: contig of 4124 bp in length  
\* 78307 78406: gap of unknown length  
\* 78407 81222: contig of 2816 bp in length  
\* 81223 81322: gap of unknown length  
\* 81323 84552: contig of 3230 bp in length  
\* 84553 84652: gap of unknown length  
\* 84653 88839: contig of 4187 bp in length  
\* 88840 88939: gap of unknown length  
\* 88940 92256: contig of 3317 bp in length  
\* 92257 92356: gap of unknown length  
\* 92357 94995: contig of 2639 bp in length  
\* 94996 95095: gap of unknown length  
\* 95096 98465: contig of 3370 bp in length  
\* 98466 98565: gap of unknown length  
\* 98566 101632: contig of 3067 bp in length  
\* 101633 101732: gap of unknown length  
\* 101733 104939: contig of 3207 bp in length  
\* 104940 105039: gap of unknown length  
\* 105040 109248: contig of 4209 bp in length  
\* 109249 109348: gap of unknown length  
\* 109349 112435: contig of 3087 bp in length  
\* 112436 112535: gap of unknown length  
\* 112536 114676: contig of 2141 bp in length  
\* 114677 114776: gap of unknown length  
\* 114777 117567: contig of 2791 bp in length  
\* 117568 117667: gap of unknown length

\* 117668 119209: contig of 1542 bp in length  
\* 119210 119309: gap of unknown length  
\* 119310 121797: contig of 2488 bp in length  
\* 121798 121897: gap of unknown length  
\* 121898 124293: contig of 2396 bp in length  
\* 124294 124393: gap of unknown length  
\* 124394 126428: contig of 2035 bp in length  
\* 126429 126528: gap of unknown length  
\* 126529 129525: contig of 2957 bp in length  
\* 129526 129626 129625: gap of unknown length  
\* 129626 132334: contig of 2709 bp in length  
\* 132335 132434: gap of unknown length  
\* 132435 135274: contig of 2840 bp in length  
\* 135275 135374: gap of unknown length  
\* 135375 138074: contig of 2700 bp in length  
\* 138075 138174: gap of unknown length  
\* 138175 139885: contig of 1811 bp in length  
\* 139886 140085: gap of unknown length  
\* 140086 142273: contig of 2188 bp in length  
\* 142274 142373: gap of unknown length  
\* 142374 143598: contig of 1225 bp in length  
\* 143599 143698: gap of unknown length  
\* 143699 145434: contig of 1736 bp in length  
\* 145435 145534: gap of unknown length  
\* 145535 146985: contig of 1451 bp in length  
\* 146986 147085: gap of unknown length  
\* 147086 148099: contig of 1014 bp in length  
\* 148100 148199: gap of unknown length  
\* 148200 150915: contig of 2716 bp in length  
\* 150916 151015: gap of unknown length  
\* 151016 152501: contig of 1486 bp in length  
\* 152502 152601: gap of unknown length  
\* 152602 154010: contig of 1409 bp in length  
\* 154011 154110: gap of unknown length  
\* 154111 155758: contig of 1648 bp in length  
\* 155759 155858: gap of unknown length  
\* 155859 157622: contig of 1764 bp in length  
\* 157623 157722: gap of unknown length  
\* 157723 159428: contig of 1706 bp in length  
\* 159429 159528: gap of unknown length  
\* 159529 161209: contig of 1681 bp in length  
\* 161210 161309: gap of unknown length  
\* 161310 163413: contig of 2104 bp in length  
\* 163414 163513: gap of unknown length  
\* 163514 164702: contig of 1189 bp in length  
\* 164703 164802: gap of unknown length  
\* 164803 165998: contig of 1196 bp in length  
\* 165999 166098: gap of unknown length  
\* 166099 167412: contig of 1314 bp in length  
\* 167413 167512: gap of unknown length  
\* 167513 169231: contig of 1719 bp in length  
\* 169232 169331: gap of unknown length  
\* 169332 170634: contig of 1203 bp in length  
\* 170635 170633: gap of unknown length  
\* 170634 172047: contig of 1413 bp in length  
\* 172048 172147: gap of unknown length  
\* 172148 173509: contig of 1362 bp in length  
\* 173510 173610 173609: gap of unknown length  
\* 173610 174953: contig of 1344 bp in length.

## FEATURES

Location/Qualifiers

Query Match 100.0%; Score 20; DB 2: Length 174953;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 20

|||||

Db 40140 GCTGGAGGATTCCTCTGTC 40121

## RESULT 4

AC121182/c

LOCUS

AC121182

157126 bp

DNA

linear

HTG 23-JUL-2002

## DEFINITION

Rattus norvegicus clone CH230-350A13, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 45 unordered pieces.

## ACCESSION

AC1211182

## VERSION

AC1211182.2 GI:21902625

## KEYWORDS

HTG: HTGS\_PHASE1.

## SOURCE

Rattus norvegicus

## ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

## REFERENCE

1 (bases 1 to 157126)

## AUTHORS

Muzay,D.M., Adams,C., Adlo-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,L.C., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbra,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Honsi,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Kamei,J., Kavar,C., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metaker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savory,G., Scher,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.

## TITLE

Direct Submission

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 157126)

## AUTHORS

Worley,K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (16-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

3 (bases 1 to 157126)

## AUTHORS

Worley,K.C.

## TITLE

Direct Submission

## JOURNAL

Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Jul 18, 2002 this sequence version replaced gi:20806209.

## JOURNAL

----- Genome Center

## AUTHORS

Center: Baylor College of Medicine

## TITLE

Center code: BCM

## JOURNAL

Web site: <http://www.hgsc.bcm.tmc.edu/>

## COMMENT

Contact: hgsc-help@bcm.tmc.edu

## JOURNAL

----- Project Information

## AUTHORS

Center project name: GWNE

## TITLE

-----

## JOURNAL

-----

## COMMENT

-----

## JOURNAL

-----

## AUTHORS

-----

## TITLE

-----

## JOURNAL

-----

## COMMENT

-----

Center clone name: CH230-350A13

----- Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; version 0.990329

Assembly program: Phrap; version 0.990329

Consensus quality: 115827 bases at least Q40

Consensus quality: 120596 bases at least Q30

Consensus quality: 124293 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 45 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

-----

\* 1 1052: contig of 1052 bp in length

\* 1053 1152: gap of unknown length

\* 1153 2472: contig of 1320 bp in length

\* 2473 2572: gap of unknown length

\* 2573 4298: contig of 1726 bp in length

\* 4299 4398: gap of unknown length

\* 4399 5674: contig of 1275 bp in length

\* 5674 5773: gap of unknown length

\* 5774 6831: contig of 1058 bp in length

\* 6832 6931: gap of unknown length

\* 6932 8024: contig of 1093 bp in length

\* 8025 8125: gap of unknown length

\* 8125 9155: contig of 1030 bp in length

\* 9155 9254: gap of unknown length

\* 9255 10489: contig of 1235 bp in length

\* 10490 10589: gap of unknown length

\* 10590 11782: contig of 1193 bp in length

\* 11783 11882: gap of unknown length

\* 11883 12919: contig of 1037 bp in length

\* 12920 13019: gap of unknown length

\* 13020 14138: contig of 1119 bp in length

\* 14139 14238: gap of unknown length

\* 14239 15370: contig of 1132 bp in length

\* 15371 15470: gap of unknown length

\* 15471 16818: contig of 1348 bp in length

\* 16819 16918: gap of unknown length

\* 16919 18749: contig of 1830 bp in length

\* 18749 18849: gap of unknown length

\* 18849 20073: contig of 1225 bp in length

\* 20074 20173: gap of unknown length

\* 20174 22545: contig of 2372 bp in length

\* 22546 22645: gap of unknown length

\* 22646 25123: contig of 2478 bp in length

\* 25124 25223: gap of unknown length

\* 25224 27943: contig of 2720 bp in length

\* 27944 28043: gap of unknown length

\* 28044 30415: contig of 2372 bp in length

\* 30416 30515: gap of unknown length

\* 30516 32575: contig of 2060 bp in length

\* 32576 32675: gap of unknown length

\* 32676 33677: contig of 1002 bp in length

\* 33678 33777: gap of unknown length

\* 33778 37129: contig of 3352 bp in length

\* 37130 37229: gap of unknown length

\* 37230 39330: contig of 2101 bp in length

\* 39331 39430: gap of unknown length

\* 39431 41849: contig of 2419 bp in length

\* 41850 41949: gap of unknown length

\* 41950 44996: contig of 3047 bp in length

\* 44997 45096: gap of unknown length

\* 45097 47943: contig of 2847 bp in length

\* 47944 48043: gap of unknown length

\* 48044 51187: contig of 3144 bp in length

\* 51188 51287: gap of unknown length

\* 51288 51287: gap of unknown length

\* 51288 51287: gap of unknown length

\* 51288 51287: gap of unknown length

\* 51288 51287: gap of unknown length

\* 51288 51287: gap of unknown length

\* 51288 51287: gap of unknown length

\* 51288 54026: contig of 2739 bp in length  
\* 54027 gap of unknown length  
\* 54126: contig of 2966 bp in length  
\* 54127 gap of unknown length  
\* 57093 57192: contig of 2890 bp in length  
\* 57193 60082: contig of 2890 bp in length  
\* 60083 60182: gap of unknown length  
\* 60183 6340: contig of 3258 bp in length  
\* 6340: contig of 3258 bp in length  
\* 6341 63540: gap of unknown length  
\* 63541 67541: contig of 4001 bp in length  
\* 67542 67642: gap of unknown length  
\* 67642 67709: contig of 5068 bp in length  
\* 67709 72710: gap of unknown length  
\* 72710 72809: gap of unknown length  
\* 72809 75669: contig of 2860 bp in length  
\* 75669 75769: gap of unknown length  
\* 75769 75770: gap of unknown length  
\* 75770 79221: contig of 3451 bp in length  
\* 79221 79321: gap of unknown length  
\* 79321 83840: contig of 4520 bp in length  
\* 83841 83940: gap of unknown length  
\* 83941 88127: contig of 4187 bp in length  
\* 88128 88227: gap of unknown length  
\* 88227 90998: contig of 2771 bp in length  
\* 90998 91099: gap of unknown length  
\* 91099 97634: contig of 6536 bp in length  
\* 97635 97735: gap of unknown length  
\* 97735 104519: contig of 6785 bp in length  
\* 104520 104619: gap of unknown length  
\* 104620 110441: contig of 5822 bp in length  
\* 110442 110541: gap of unknown length  
\* 110542 121487: contig of 10946 bp in length  
\* 121488 121587: gap of unknown length  
\* 121588 131266: contig of 9879 bp in length  
\* 131267 131366: gap of unknown length  
\* 131367 142365: contig of 10999 bp in length  
\* 142366 142465: gap of unknown length  
\* 142466 157126: contig of 14661 bp in length.

Location/Qualifiers

1. 157126  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-350A13"  
BASE COUNT 41349 a 34061 c 33998 g 40715 t 7103 others

FEATURES  
source

Query Match 95.0%; Score 19; DB 2; Length 157126;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CTGGAGGATTCCTGTGCC 20  
Db 61633 CTGGAGGATTCCTGTGCC 61615

RESULT 5  
AC095165/c 179992 bp DNA linear HTG 11-JUL-2002  
LOCUS Rattus norvegicus clone CH230-8D23, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
DEFINITION 53 unordered pieces.  
AC095165  
VERSION AC095165.3 GI:21722645  
KEYWORDS HTG; HGSC\_PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 179992)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhaý,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,

Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denu,A.L., Ding,X., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulsegad,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,T., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teiford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,X., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 179992)  
Worley K.C.  
Direct Submission  
Submitted (16-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 179992)  
Worley K.C.  
Direct Submission  
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence version replaced gi:17942209.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GCXG  
Center clone name: CH230-8D23  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 141474 bases at least Q40  
Consensus quality: 147236 bases at least Q30  
Consensus quality: 151798 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 53 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will

\* be preserved.

1206: contig of 1206 bp in length  
1207: gap of unknown length  
1306: contig of 1243 bp in length  
2549: gap of unknown length  
2649: gap of unknown length  
4153: contig of 1504 bp in length  
4253: gap of unknown length  
5378: contig of 1125 bp in length  
5478: gap of unknown length  
5541: contig of 1063 bp in length  
6641: gap of unknown length  
7646: contig of 1005 bp in length  
7746: gap of unknown length  
8961: contig of 1215 bp in length  
9061: gap of unknown length  
10473: contig of 1412 bp in length  
10573: gap of unknown length  
12144: contig of 1571 bp in length  
12244: gap of unknown length  
14540: contig of 2296 bp in length  
14640: gap of unknown length  
15971: contig of 1331 bp in length  
16071: gap of unknown length  
17346: contig of 1275 bp in length  
17446: gap of unknown length  
19676: contig of 2230 bp in length  
19776: gap of unknown length  
21890: contig of 2114 bp in length  
21990: gap of unknown length  
23848: contig of 1858 bp in length  
23948: gap of unknown length  
25759: contig of 1811 bp in length  
25859: gap of unknown length  
27618: contig of 1759 bp in length  
27718: gap of unknown length  
29318: contig of 1600 bp in length  
29319: gap of unknown length  
30580: contig of 1162 bp in length  
30680: gap of unknown length  
33145: contig of 2465 bp in length  
33245: gap of unknown length  
35872: contig of 2627 bp in length  
35972: gap of unknown length  
37797: contig of 1825 bp in length  
37897: gap of unknown length  
40940: contig of 3043 bp in length  
41040: gap of unknown length  
42355: contig of 1315 bp in length  
42455: gap of unknown length  
44238: contig of 1783 bp in length  
44338: gap of unknown length  
47040: contig of 2702 bp in length  
47140: gap of unknown length  
50437: contig of 3297 bp in length  
50537: gap of unknown length  
53024: contig of 2487 bp in length  
53124: gap of unknown length  
53284: contig of 3160 bp in length  
56384: gap of unknown length  
60012: contig of 3628 bp in length  
60013: gap of unknown length  
63050: contig of 2938 bp in length  
63150: gap of unknown length  
67104: contig of 3954 bp in length  
67204: gap of unknown length  
69870: contig of 2666 bp in length  
69970: gap of unknown length  
73460: contig of 3490 bp in length  
73560: gap of unknown length  
75659: contig of 2099 bp in length  
75759: gap of unknown length  
78360: contig of 2601 bp in length  
78460: gap of unknown length

78461 81537: contig of 3077 bp in length  
81538 81637: gap of unknown length  
81638 85252: contig of 3615 bp in length  
85253 85352: gap of unknown length  
85353 88826: contig of 3474 bp in length  
88827 88926: gap of unknown length  
88927 92479: contig of 3553 bp in length  
92480 92579: gap of unknown length  
92580 97441: contig of 4862 bp in length  
97442 97541: gap of unknown length  
97542 101258: contig of 3717 bp in length  
101259 101358: gap of unknown length  
101359 106514: contig of 5156 bp in length  
106515 106615: gap of unknown length  
110465 110466: gap of unknown length  
110565 110566: gap of unknown length  
116831 116930: contig of 6265 bp in length  
122453 122452: contig of 5522 bp in length  
122552 122553: gap of unknown length  
127781 127780: contig of 5228 bp in length  
127881 127880: gap of unknown length  
135358 135357: gap of unknown length  
135458 142962: contig of 7505 bp in length  
142963 143062: gap of unknown length  
143063 151330: contig of 8268 bp in length  
151331 151430: gap of unknown length  
151431 160041: contig of 8611 bp in length  
160042 160141: gap of unknown length  
160142 166777: contig of 6636 bp in length  
166778 166877: gap of unknown length  
166878 179992: contig of 13115 bp in length.

FEATURES  
source Location/Qualifiers  
1. 179992

Query Match 95.0%; Score 19; DB 2; Length 179992;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTGGAGGATTCCTCTGTC 20  
|||||

Db 167036 CTGGAGGATTCCTCTGTC 167018

RESULT 6

MUSKSD3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

MUSKSD3 1592 bp mRNA linear ROD 17-OCT-1997  
Mus musculus 3-ketosteroid reductase (HSD3b5) mRNA, complete cds.  
L41519  
L41519.1 GI:840649

Mus musculus (strain BALB/c, sub-species domesticus) (clone  
library: lambda gtl) male adult liver cDNA to mRNA.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 1592)

Abbaszade, I.G., Clarke, T.R., Park, C.H. and Payne, A.H.  
The mouse 3 beta-hydroxysteroid dehydrogenase multigene family  
includes two functionally distinct groups of proteins  
Mol. Endocrinol. 9 (9), 1214-1222 (1995)

7491113

2 (bases 1 to 1592)

Payne, A.

Direct Submission

Submitted (17-OCT-1997) Gyn/Ob, Stanford University Medical Center,

300 Pasteur Dr., Stanford, CA 94305, USA

Location/Qualifiers

1. 1592

/organism="Mus musculus"

/strain="BALB/c"

```

/sub_species="domesticus"
/db_xref="taxon:10090"
/sex="male"
/tissue_type="liver"
/clone_lib="lambda gt11"
/dev_stage="adult"
1. 1592
/gene="HSD3b5"
100. 1221
/gene="HSD3b5"
/note="3-beta-hydroxysteroid dehydrogenase;
3-hydroxysteroid; ketosteroid; multigene family"
/codon_start=1
/evidence-experimental
/product="3-ketosteroid reductase"
/protein_id="AAB81242.1"
/db_xref="GI:840650"
/translation="MPGWSCLVTGAGFLGQIRVLMVOEELQRIALFRFFGRKHE
BELSKLOPKAKVRLKGLDILDAQCLKACQGNMSAVIHTAAADPRGAASROPIIDVNL
KGTQLLDACVEASVPTFIYSSVLVAGPNSYKEIILNAHEEHESHSTWPNPYPSKR
MAEKAVLATNGRLKANGTLHTCALRPLFIYGEEOVSTTVKTKNNISIKKNATF
SIANPVYGNAAWAHILAAISLDQPKSPSQIQQFYISDNTPHQSYDLDLNTLSKEW
GLDLSGWRLPLSLILYLAFLLETVSLFRLPVYVNPPTRLTLITVLNSVFIYSYKKA
ORDLGYPILVSWEEAKQKTSIEWIGLVKQHRETLHKKSQ"
1573. 1579
/gene="HSD3b5"
1592
/gene="HSD3b5"
431 a 404 c 371 g 386 t
BASE COUNT
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 1592;
Best Local Similarity 95.0%; Pred. NO. 18;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTGTGCC 20
Db 947 GCTGGAGGATTCCTGTGCC 966

RESULT 7
AC022586
LOCUS
DEFINITION
Homo sapiens chromosome 2 clone RP11-388J8 map 2, LOW-PASS SEQUENCE
SAMPLING.
AC022586
VERSION
AC022586.2 GI:9123913
KEYWORDS
HTG; HTGS_PHASE0.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 77333)
Birren.B., Linton.L., Nusbaum.C. and Lander.E.
Homo sapiens chromosome 2, clone RP11-388J8
Unpublished
REFERENCE
2 (bases 1 to 77333)
Birren.B., Linton.L., Nusbaum.C., Lander.E., Abraham.H., Allen.N.,
Anderson.S., Baldwin.J., Barna.N., Beckerly.R., Bede.F.,
Boguslavsky.L., Bouckgalter.B., Brown.A., Burkett.G., Castle.A.,
Choepel.Y., Collangelo.M., Collins.S., Collamore.A., Cooke.P.,
Dearellano.K., Dewar.K., Domino.M., Doyle.M., Fenestor.J.,
Ferreira.P., FitzHugh.W., Forrest.C., Gage.D., Galagan.J.,
Gardyna.S., Grant.G., Hagos.B., Heaford.A., Horton.L.,
Howland.J.C., Johnson.R., Jones.C., Kann.L., Karatas.A., Klein.J.,
Landers.T., Lechoczky.J., Levine.R., Liu.C., Liu.G., Locke.K.,
Macdonald.P., Marquis.N., McEwan.P., McGurk.A., McKernan.K.,
McPheeters.R., Meldrum.J., Meneus.L., Morrow.J., Naylor.J.,
Norman.C.H., O'Connor.T., O'Donnell.P., Oliver.T.M., Peterson.K.,
Pierre.N., Pisani.C., Pollara.V., Raymond.C., Riley.R., Rothman.D.,
Roy.A., Santos.R., Severi.P., Spencer.B., Stange-Thomann.N.,
Stojanovic.N., Subramanian.A., Talamas.J., Tesfaye.S., Theodore.J.,
Tirrell.A., Vassiliev.H., Viel.R., Vo.A., Wu.X., Wyman.D., Ye.W.J.,

```

# TITLE JOURNAL COMMENT

Zimmer A. and Zody, M.  
Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 13, 2000 this sequence version replaced gi:6910752.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L3786  
Center clone name: 388\_J-8  
-----

\* NOTE: This record contains 78 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 846: contig of 846 bp in length  
\* 847 946: gap of 100 bp  
\* 947 1849: contig of 903 bp in length  
\* 1850 1949: gap of 100 bp  
\* 1950 2837: contig of 888 bp in length  
\* 2838 2937: gap of 100 bp  
\* 2938 3841: contig of 904 bp in length  
\* 3842 3941: gap of 100 bp  
\* 3942 4787: contig of 846 bp in length  
\* 4788 4887: gap of 100 bp  
\* 4888 5797: contig of 910 bp in length  
\* 5798 5897: gap of 100 bp  
\* 5898 6804: contig of 907 bp in length  
\* 6805 6904: gap of 100 bp  
\* 6905 7785: contig of 881 bp in length  
\* 7786 7885: gap of 100 bp  
\* 7886 8776: contig of 891 bp in length  
\* 8777 8876: gap of 100 bp  
\* 8778 9734: contig of 858 bp in length  
\* 9735 9834: gap of 100 bp  
\* 9835 10743: contig of 908 bp in length  
\* 10743 10842: gap of 100 bp  
\* 10843 11746: contig of 904 bp in length  
\* 11747 11846: gap of 100 bp  
\* 11847 12699: contig of 853 bp in length  
\* 12700 12799: gap of 100 bp  
\* 12800 13704: contig of 905 bp in length  
\* 13705 13804: gap of 100 bp  
\* 13805 14701: contig of 897 bp in length  
\* 14702 14801: gap of 100 bp  
\* 14802 15715: contig of 914 bp in length  
\* 15716 15815: gap of 100 bp  
\* 15816 16720: contig of 905 bp in length  
\* 16721 16820: gap of 100 bp  
\* 16821 17672: contig of 852 bp in length  
\* 17673 17772: gap of 100 bp  
\* 17773 18721: contig of 949 bp in length  
\* 18722 18821: gap of 100 bp  
\* 18822 19690: contig of 869 bp in length  
\* 19691 19790: gap of 100 bp  
\* 19791 20676: contig of 886 bp in length  
\* 20677 20776: gap of 100 bp  
\* 20777 21659: contig of 883 bp in length  
\* 21660 21759: gap of 100 bp  
\* 21760 22655: contig of 896 bp in length

22656 22755: gap of 100 bp  
 22756 23657: contig of 902 bp in length  
 23658 23757: gap of 100 bp  
 23758 24633: contig of 876 bp in length  
 24634 24733: gap of 100 bp  
 24734 25603: contig of 876 bp in length  
 25610 25709: gap of 100 bp  
 25710 26614: contig of 905 bp in length  
 26615 26714: gap of 100 bp  
 26715 27595: contig of 881 bp in length  
 27596 27695: gap of 100 bp  
 27696 28596: contig of 901 bp in length  
 28597 28696: gap of 100 bp  
 28697 29563: contig of 867 bp in length  
 29564 29663: gap of 100 bp  
 29664 30541: contig of 878 bp in length  
 30542 30641: gap of 100 bp  
 30642 31543: contig of 902 bp in length  
 31544 31643: gap of 100 bp  
 31644 32534: contig of 891 bp in length  
 32535 32634: gap of 100 bp  
 32635 33566: contig of 932 bp in length  
 33567 33666: gap of 100 bp  
 33667 34578: contig of 912 bp in length  
 34579 34678: gap of 100 bp  
 34679 35567: contig of 889 bp in length  
 35568 35667: gap of 100 bp  
 35668 36536: contig of 869 bp in length  
 36537 36636: gap of 100 bp  
 36637 37544: contig of 908 bp in length  
 37545 37644: gap of 100 bp  
 37645 38538: contig of 894 bp in length  
 38539 38638: gap of 100 bp  
 38639 39502: contig of 864 bp in length  
 39503 39602: gap of 100 bp  
 39603 40487: contig of 885 bp in length  
 40488 40587: gap of 100 bp  
 40588 41438: contig of 851 bp in length  
 41439 41538: gap of 100 bp  
 41539 42435: contig of 897 bp in length  
 42436 42535: gap of 100 bp  
 42536 43474: contig of 939 bp in length  
 43475 43574: gap of 100 bp  
 43575 44424: contig of 850 bp in length  
 44425 44524: gap of 100 bp  
 44525 45409: contig of 885 bp in length  
 45410 45509: gap of 100 bp  
 45510 46403: contig of 894 bp in length  
 46404 46503: gap of 100 bp  
 46504 47407: contig of 904 bp in length  
 47408 47507: gap of 100 bp  
 47508 48400: contig of 893 bp in length  
 48401 48500: gap of 100 bp  
 48501 49371: contig of 871 bp in length  
 49372 49471: gap of 100 bp  
 49472 50347: contig of 876 bp in length  
 50348 50447: gap of 100 bp  
 50448 51370: contig of 923 bp in length  
 51371 51470: gap of 100 bp  
 51471 52361: contig of 891 bp in length  
 52362 52461: gap of 100 bp  
 52462 53377: contig of 916 bp in length  
 53378 53477: gap of 100 bp  
 53478 54350: contig of 873 bp in length  
 54351 54450: gap of 100 bp  
 54451 55349: contig of 899 bp in length  
 55350 55449: gap of 100 bp  
 55450 56328: contig of 879 bp in length  
 56329 56428: gap of 100 bp  
 56429 57320: contig of 892 bp in length  
 57321 57420: gap of 100 bp  
 57421 58287: contig of 867 bp in length  
 58288 58387: gap of 100 bp

58388 59299: contig of 912 bp in length  
 59300 59399: gap of 100 bp  
 59400 60286: contig of 887 bp in length  
 60287 60386: gap of 100 bp  
 60387 61273: contig of 887 bp in length  
 61274 61373: gap of 100 bp  
 61374 62291: contig of 918 bp in length  
 62292 62391: gap of 100 bp  
 62392 63339: contig of 948 bp in length  
 63340 63439: gap of 100 bp  
 63440 64333: contig of 894 bp in length  
 64334 64433: gap of 100 bp  
 64434 65323: contig of 890 bp in length  
 65324 65423: gap of 100 bp  
 65424 66327: contig of 904 bp in length  
 66328 66427: gap of 100 bp  
 66428 67304: contig of 877 bp in length  
 67305 67404: gap of 100 bp  
 67405 68292: contig of 888 bp in length  
 68293 68392: gap of 100 bp  
 68393 69303: contig of 911 bp in length  
 69304 69403: gap of 100 bp  
 69404 70301: contig of 898 bp in length

Query Match 92.0%; Score 18.4; DB 2; Length 77333;  
 Best Local Similarity 95.0%; Pred. No. 23;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTCC 20  
 Db 24111 GCTGGAGGCTTCCTGTCC 24130

## RESULT 8

AC009505 AC009505 163880 bp DNA linear PRI 30-SEP-2000  
 LOCUS Homo sapiens BAC clone RP11-526D2 from 2, complete sequence.  
 DEFINITION AC009505  
 ACCESSION AC009505  
 VERSION AC009505.3 GI:8748933  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 163880)  
 AUTHORS Sulston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 163880)  
 AUTHORS Armstrong, J., Maupin, R., Haakenson, W. and Glaser, E.  
 TITLE The sequence of Homo sapiens BAC clone RP11-526D2  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 163880)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-AUG-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 163880)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-JUN-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 5 (bases 1 to 163880)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Jun 27, 2000 this sequence version replaced gi:6042117.  
 ----- Genome Center



```

misc_feature 26295..26791 /note="similar to EST AW912919 (NID:g8078549) uf46e08.y1"
misc_feature 26310..26612 /note="similar to EST AA722674 (NID:g2740381) zg90h06.sl"
misc_feature 26311..26747 /note="similar to EST AA133513 (NID:g1690500) zk96g02.rl"
misc_feature 26311..26652 /note="similar to EST AA036848 (NID:g1509886) zk29f08.rl"
misc_feature 26317..26940 /note="similar to EST AI868961 (NID:g5542929) wc17ell.x1"
misc_feature 26353..26942 /note="similar to EST AI934552 (NID:g5673349) wp09c10.x1"
misc_feature 26367..26940 /note="similar to EST AA115694 (NID:g1670786) zk96g02.sl"
misc_feature 26374..26689 /note="similar to EST AW371193 (NID:g6875847)"

Query Match 92.0%; Score 18.4; DB 9; Length 163880;
Best Local Similarity 95.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GCTGGAGGATTCCTCTGTCC 20
      ||||| ||||| ||||| ||||| |||||
Db 50195 GCTGGAGGCTTCTCTGTCC 50214

RESULT 9
AL627184
LOCUS
DEFINITION Mus musculus chromosome 4 clone RP23-125F21, linear HTG 29-JUN-2002
ACCESSION AL627184
VERSION 214433 bp DNA linear HTG 29-JUN-2002
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 214433)
JOURNAL Blakey, S.

```

```

COMMENT
Submitted (28-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquies@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 2, 2002 this sequence version replaced gi:21627911.
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquies@sanger.ac.uk
Project Information
Center project name: BM125F21
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: pBlasmid; L08752; 100% of reads
Chemistry: Dye-terminator; 6% of reads
Chemistry: Dye-terminator Big Dye; 93% of reads
Consensus quality: 214401 bases at least Q40
Consensus quality: 214426 bases at least Q30
Consensus quality: 214429 bases at least Q20
Insert size: 200937; 13.8% error; agarose-ff
Quality coverage: 14.57x in Q20 bases; sum-of-contigs Quality
coverage: 15.70x in Q20 bases; agarose-ff

```

```

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..214433
/organism="Mus musculus"

```

```

FEATURES
source

```

```

/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-125F21"
/clone.lib="RPCI-23"
1..214433
/note="assembly_fragment:05960
clone_end:SP6
vector_side:left"
BASE COUNT 52970 a 53858 c 55148 g 52457 t
ORIGIN

```

```

Query Match 92.0%; Score 18.4; DB 2; Length 214433;
Best Local Similarity 95.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCTGGAGGATTCCTCTGTCC 20
      ||||| ||||| ||||| ||||| |||||
Db 106804 GCTGGAGGATTCCTCTGTCC 106823

```

```

RESULT 10
AC095501
LOCUS
DEFINITION Rattus norvegicus clone CH230-7P13, linear HTG 09-JUL-2002
ACCESSION AC095501
VERSION 110491 bp DNA linear HTG 09-JUL-2002
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 110491)

```

```

REFERENCE
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbata, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Delaney, K.R., Delgado, O., Dunn, A.L., Dugan-Rocha, S., Durbin, K.J.,
Douthwaite, K.J., Draper, H., Edwards, C.C., Elhaj, C., Escotto, M.,
Earhart, C., Edgar, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,
Karlsone, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,
Lozad, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G.,
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
Rives, M., Rojas, A., Rojokob, I., Rolfe, M., Ruiz, S., Savary, G.,
Scherefer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,

```

Weinstock, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 110491)  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

Direct Submission  
 Submitted (09-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 9, 2002 this sequence version replaced gi:17942008.

COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GCME  
 Center clone name: CH230-7P13  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 47586 bases at least Q40  
 Consensus quality: 51558 bases at least Q30  
 Consensus quality: 54897 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 61 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1014: contig of 1014 bp in length  
 1015 1114: gap of unknown length  
 1115 2127: contig of 1013 bp in length  
 2128 2227: gap of unknown length  
 2228 3250: contig of 1023 bp in length  
 3251 3351: gap of unknown length  
 3352 4354: contig of 1004 bp in length  
 4355 4454: gap of unknown length  
 4455 6134: contig of 1680 bp in length  
 6135 6234: gap of unknown length  
 6235 7339: contig of 1105 bp in length  
 7340 7439: gap of unknown length  
 7440 8767: contig of 1328 bp in length  
 8768 8867: gap of unknown length  
 8868 9891: contig of 1024 bp in length  
 9892 9991: gap of unknown length  
 9992 11850: contig of 1859 bp in length  
 11851 11950: gap of unknown length  
 11951 13496: contig of 1546 bp in length  
 13497 13596: gap of unknown length  
 13597 14746: contig of 1150 bp in length  
 14747 14846: gap of unknown length  
 14847 15972: contig of 1126 bp in length  
 15973 16072: gap of unknown length  
 16073 17477: contig of 1405 bp in length  
 17478 17577: gap of unknown length  
 17578 18598: contig of 1021 bp in length  
 18599 18698: gap of unknown length  
 18699 19843: contig of 1145 bp in length  
 19844 20998: gap of unknown length  
 20998 20998: contig of 1055 bp in length

20999 21098: gap of unknown length  
 21099 22367: contig of 1269 bp in length  
 22368 22467: gap of unknown length  
 22468 23904: contig of 1437 bp in length  
 23905 24004: gap of unknown length  
 24005 25186: contig of 1182 bp in length  
 25187 25286: gap of unknown length  
 25287 26385: contig of 1099 bp in length  
 26386 26485: gap of unknown length  
 26486 28106: contig of 1621 bp in length  
 28107 28206: gap of unknown length  
 28207 29694: contig of 1488 bp in length  
 29695 30895: contig of 1101 bp in length  
 30896 32160: contig of 1165 bp in length  
 32161 32260: gap of unknown length  
 32261 33874: contig of 1614 bp in length  
 33875 35528: contig of 1554 bp in length  
 35529 37052: contig of 1424 bp in length  
 37053 37152: gap of unknown length  
 37153 38831: contig of 1679 bp in length  
 38832 40487: contig of 1556 bp in length  
 40488 40587: gap of unknown length  
 40588 42872: contig of 2285 bp in length  
 42873 42972: gap of unknown length  
 42973 44569: contig of 1597 bp in length  
 44570 46570: gap of unknown length  
 46571 46670: contig of 1901 bp in length  
 46671 48333: contig of 1663 bp in length  
 48334 48434: gap of unknown length  
 48434 49766: contig of 1333 bp in length  
 49767 52014: contig of 2148 bp in length  
 52015 52114: gap of unknown length  
 52115 53981: contig of 1867 bp in length  
 53982 54081: gap of unknown length  
 54082 55632: contig of 1551 bp in length  
 55633 57322: gap of unknown length  
 57323 57460: contig of 1728 bp in length  
 57461 57560: gap of unknown length  
 57561 59263: contig of 1703 bp in length  
 59264 59363: gap of unknown length  
 59364 60449: contig of 1086 bp in length  
 60450 62711: contig of 2162 bp in length  
 62712 62811: gap of unknown length  
 62812 64168: contig of 1357 bp in length  
 64169 64268: gap of unknown length  
 64269 65460: contig of 1192 bp in length  
 65461 65560: gap of unknown length  
 65561 67493: contig of 1933 bp in length  
 67494 67593: gap of unknown length  
 67594 70085: contig of 2492 bp in length  
 70086 70185: gap of unknown length  
 70186 72041: contig of 1856 bp in length  
 72042 72141: gap of unknown length  
 72142 74011: contig of 1870 bp in length  
 74012 74111: gap of unknown length  
 74112 76007: contig of 1896 bp in length  
 76008 76107: gap of unknown length  
 76108 78585: contig of 2478 bp in length  
 78586 78686: gap of unknown length  
 78686 81186: contig of 2501 bp in length  
 81187 81286: gap of unknown length  
 81287 82810: contig of 1524 bp in length  
 82811 82911: gap of unknown length  
 82911 85151: contig of 2241 bp in length  
 85152 85251: gap of unknown length

\* 85252 87452: contig of 2201 bp in length  
 \* 87552: gap of unknown length  
 \* 87553 90136: contig of 2584 bp in length  
 Query Match 90.0%; Score 18; DB 2; Length 110491;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGT 18  
 Db 2482 GCTGGAGGATTCCTCTGT 2499  
 RESULT 11  
 AC095500/c  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-7p12, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 ACCESSION AC095500  
 VERSION 71 unordered pieces.  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 158869)  
 REFERENCE  
 AUTHORS Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Dena,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gumarate,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Lucier,A., Lucia,J., Liu,W., Loulseged,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokweso,S., Oguh,M., Okwuonu,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Rives,M., Rojals,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,  
 Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Woodson,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 158869)  
 REFERENCE  
 AUTHORS Worley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 158869)  
 Worley,K.C.  
 Direct Submission  
 Submitted (10-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 9, 2002 this sequence version replaced gi:17942007.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GCMC  
 Center clone name: CH230-7P12  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 88636 bases at least Q40  
 Consensus quality: 95512 bases at least Q30  
 Consensus quality: 100849 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 71 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1494: contig of 1494 bp in length  
 1495 1594: gap of unknown length  
 1595 2796: contig of 1202 bp in length  
 2797 2896: gap of unknown length  
 2897 4154: contig of 1258 bp in length  
 4155 4254: gap of unknown length  
 4255 5337: contig of 1083 bp in length  
 5338 5437: gap of unknown length  
 5438 6490: contig of 1053 bp in length  
 6491 6591: gap of unknown length  
 6591 7618: contig of 1027 bp in length  
 7618 7717: gap of unknown length  
 7718 8995: contig of 1278 bp in length  
 8996 9095: gap of unknown length  
 9096 10242: contig of 1147 bp in length  
 10243 10343: contig of unknown length  
 10343 11631: gap of unknown length  
 11631 11731: gap of unknown length  
 11731 12968: contig of 1238 bp in length  
 12969 13068: gap of unknown length  
 13069 14567: contig of 1499 bp in length  
 14568 14667: gap of unknown length  
 14668 15918: contig of 1251 bp in length  
 15919 16019: gap of unknown length  
 16019 17338: contig of 1320 bp in length  
 17339 17438: gap of unknown length  
 17439 18803: contig of 1365 bp in length  
 18804 18903: gap of unknown length  
 18904 19977: contig of 1074 bp in length  
 19978 20077: gap of unknown length  
 20078 21226: contig of 1149 bp in length  
 21227 21326: gap of unknown length  
 21327 22851: contig of 1525 bp in length  
 22852 22951: gap of unknown length  
 22952 24497: contig of 1546 bp in length  
 24498 24597: gap of unknown length  
 24598 26328: contig of 1731 bp in length  
 26329 26428: gap of unknown length

\* 26429 contig of 1445 bp in length  
\* 27873: gap of unknown length  
\* 27874 contig of 1107 bp in length  
\* 27974 contig of 1107 bp in length  
\* 29080: contig of 1036 bp in length  
\* 29180: contig of 1036 bp in length  
\* 30216: contig of 1168 bp in length  
\* 30317: contig of 1168 bp in length  
\* 31484: contig of 1168 bp in length  
\* 31485: gap of unknown length  
\* 31585: contig of 1173 bp in length  
\* 32757: gap of unknown length  
\* 32857: gap of unknown length  
\* 32858: contig of 1659 bp in length  
\* 34517: gap of unknown length  
\* 34616: gap of unknown length  
\* 35906: contig of 1289 bp in length  
\* 36006: gap of unknown length  
\* 36006: gap of unknown length  
\* 37093: contig of 1088 bp in length  
\* 37193: gap of unknown length  
\* 37194: contig of 1360 bp in length  
\* 38553: gap of unknown length  
\* 38554: contig of 1072 bp in length  
\* 38654: contig of 1072 bp in length  
\* 39726: gap of unknown length  
\* 39826: contig of 1717 bp in length  
\* 41543: gap of unknown length  
\* 41642: gap of unknown length  
\* 41643: contig of 1261 bp in length  
\* 42904: gap of unknown length  
\* 43003: gap of unknown length  
\* 43004: contig of 2007 bp in length  
\* 45010: gap of unknown length  
\* 45110: gap of unknown length  
\* 45111: contig of 1498 bp in length  
\* 46609: gap of unknown length  
\* 46709: contig of 1504 bp in length  
\* 48212: gap of unknown length  
\* 48213: gap of unknown length  
\* 48313: contig of 1812 bp in length  
\* 50125: gap of unknown length  
\* 50224: gap of unknown length  
\* 50225: contig of 1267 bp in length  
\* 51492: gap of unknown length  
\* 51492: gap of unknown length  
\* 51591: gap of unknown length  
\* 51592: contig of 1582 bp in length  
\* 53174: gap of unknown length  
\* 53273: gap of unknown length  
\* 53274: contig of 1853 bp in length  
\* 55127: gap of unknown length  
\* 55227: contig of 1752 bp in length  
\* 56978: gap of unknown length  
\* 56979: gap of unknown length  
\* 57079: contig of 1939 bp in length  
\* 59018: gap of unknown length  
\* 59117: gap of unknown length  
\* 60347: contig of 1230 bp in length  
\* 60348: gap of unknown length  
\* 60447: gap of unknown length  
\* 62287: contig of 1840 bp in length  
\* 62388: gap of unknown length  
\* 62388: gap of unknown length  
\* 62388: gap of unknown length  
\* 64079: contig of 1692 bp in length  
\* 64080: gap of unknown length  
\* 64179: gap of unknown length  
\* 64180: contig of 2314 bp in length  
\* 64180: gap of unknown length  
\* 64994: gap of unknown length  
\* 65994: contig of 2440 bp in length  
\* 65994: gap of unknown length  
\* 69033: gap of unknown length  
\* 69034: gap of unknown length  
\* 69134: contig of 1707 bp in length  
\* 70841: gap of unknown length  
\* 70841: gap of unknown length  
\* 70941: contig of 1717 bp in length  
\* 72658: gap of unknown length  
\* 72757: gap of unknown length  
\* 72758: contig of 2483 bp in length  
\* 75241: gap of unknown length  
\* 75241: gap of unknown length  
\* 75341: contig of 1952 bp in length  
\* 77293: gap of unknown length  
\* 77293: gap of unknown length  
\* 77393: contig of 1517 bp in length  
\* 78909: gap of unknown length  
\* 78910: contig of 2760 bp in length  
\* 81769: gap of unknown length  
\* 81770: gap of unknown length  
\* 81870: contig of 2697 bp in length  
\* 84566: gap of unknown length  
\* 84567: gap of unknown length  
\* 84667: contig of 2466 bp in length  
\* 87133: gap of unknown length  
\* 87133: gap of unknown length  
\* 89015: contig of 1783 bp in length

Query Match 90.0%; Score 18; DB 2; Length 158869;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTGGAGGATTCCTCTGT 18  
Db 119728 GCTGGAGGATTCCTCTGT 119711  
RESULT 12  
AY058691/c 2773 bp mRNA linear INV 17-OCT-2001  
LOCUS Drosophila melanogaster LD41005 full length cDNA.  
DEFINITION AY058691  
ACCESSION AY058691  
VERSION FLI.CDNA.  
KEYWORDS Drosophila melanogaster.  
SOURCE Drosophila melanogaster.  
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 2773)  
AUTHORS Stapleton, M., Brokstein, P., Hong, L., Agbayani, A., Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George, R., Gonzalez, M., Guarin, H., Li, P., Liao, G., Miranda, A., Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Rubin, G. M. and Celniker, S.  
Direct Submission  
Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

## FEATURES

Location/Qualifiers  
1..2773  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="57E6-57E8"  
1..2773  
/gene="CG10496"  
/note="alignment with genomic scaffold AE003454"  
/db\_xref="FLYBASE:FBgn0034631"  
94..2622  
/gene="CG10496"  
/note="longest ORF"  
/codon\_start=1  
/product="LD41005p"  
/protein\_id="AAL13920.1"  
/db\_xref="GI:16198211"  
/db\_xref="FLYBASE:FBgn0034631"  
/translation="MPRTYOTALNSAKYSDQVTKTSDNESDEDDLPDQDAEDNNE  
PDLRLSILENAPHFSYPLGTMLGQLLYDQHPKODLPNCVYVPLPLFLARDGOT  
PRVEACVQPFYDDTQLOLQKALFDRRLKMRQVQEQGVHDSALFENAEHTA  
QOENFFRDSYDYTYTGNNLTIPFDGNQLAMHVSNGKRLDLHFSDYGEAFWQPL  
HSATIEESSEVFELPLHLNFRSOMFALLKLESLYELKCKEODSDSDYLIRHC  
MFSSQVAPFTSAQSLANANTLALASQDRSLRFVDISTQODIAKLDVCLLGLKTKTS  
TWAQKQAEESTFHCITQPVFTVDVRCNLPNCPFCASGVHSCACERFSCLAGVNP  
LLYASNNHNFCLDIRCLGKLTDRVAVTWTHTQMTYPPCFIDTSSVHODSEYIALGGLL

PKDLRICELOGAQKSVDMFSPALPPAPVTLLEALDARLGGFVDVYADIAERVTC  
 ITGMRITGEGASDQAFQTLLTSLGSDVYQCORLTLRDEEHEKQVQITGHTHTAIFY  
 LANAVNRVQREINQVOSIPEIRIDFRDGAQRSPKEPKLIVIEIDYIEDTD  
 VSEDESPASEKVEEPSKGGKTKPKKTKVQPKPKPKKAASRNKVKQKSISSG  
 PWOKSAYKLSHYTDMSTRLLDWMDEEDTTRDREMFEDEKIKELSPKPEGMANN  
 LEQPSOTVVPPEDETEGNDPWPQTSPLKLYGATTANYSVIPALNSDQDGERLIS  
 PKPESVPEREYSIILPQNTIIEIDYPCAPPAPKPKVKRVKGF\*  
 857 a 619 c 652 g 645 t

Query Match 87.0%; Score 17.4; DB 3; Length 2773;  
 Best Local Similarity 94.7%; Pred. No. 72;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 GCTGGAGGATTCTCTGTC 19  
 DB 777 GCTGGAGGATTCTCTATC 759

RESULT 13  
 AC095847  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-10D4, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 AC095847 2937 bp DNA linear HTG 11-JUL-2002  
 16 unordered pieces.  
 AC095847.3 GI:21722915  
 HTG: HTGS-PHASE1  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 2937)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Ayale, M., Banks, T.,  
 Barbaria, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Bulah, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, F., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, G., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, J., Liu, J., Liu, W., Louissegh, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwunodu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, Y., Villalobos, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Woodden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G. and Gibbs, R.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Direct Submission  
 Unpublished  
 2 (bases 1 to 29937)  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 29937)  
 Worley, K.C.  
 Direct Submission  
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 10, 2002 this sequence version replaced gi:17943453.  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 -----  
 Project Information  
 Center project name: GDPA  
 Center clone name: CH230-10D4  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 12551 bases at least Q40  
 Consensus quality: 13751 bases at least Q30  
 Consensus quality: 14782 bases at least Q20  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 16 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1008: contig of 1008 bp in length  
 1009 1108: gap of unknown length  
 1109 2291: contig of 1183 bp in length  
 2292 2391: gap of unknown length  
 2392 3428: contig of 1037 bp in length  
 3429 3529: gap of unknown length  
 3529 4965: contig of 1436 bp in length  
 4965 5065: gap of unknown length  
 5065 6288: contig of 1224 bp in length  
 6288 6389: gap of unknown length  
 6389 7949: contig of 1560 bp in length  
 7949 8049: gap of unknown length  
 8049 9162: contig of 1114 bp in length  
 9162 9263: gap of unknown length  
 9263 11186: contig of 1924 bp in length  
 11186 13407: gap of unknown length  
 13407 13507: gap of unknown length  
 13507 14868: contig of 1361 bp in length  
 14868 14968: gap of unknown length  
 14968 16789: contig of 1821 bp in length  
 16789 16890: gap of unknown length  
 16890 18557: contig of 1667 bp in length  
 18557 18657: gap of unknown length  
 18657 21074: contig of 2318 bp in length  
 21074 22994: gap of unknown length  
 22994 23094: contig of 1920 bp in length  
 23094 25465: gap of unknown length  
 25465 25566: contig of 2371 bp in length  
 25566 29937: gap of unknown length  
 29937 4372: contig of 4372 bp in length.

Location/Qualifiers

FEATURES

```

source
1. 29937
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-10D4"
BASE COUNT 7213 a 6807 c 6699 g 7702 t 1516 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 29937;
Best Local Similarity 94.7%; Pred. No. 85;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19
|||||
Db 29884 GCTGGAGGATTCCTCTGTC 29902

RESULT 14
AC019884/c
LOCUS
DEFINITION Drosophila melanogaster, 54186 bp DNA linear HTG 03-JAN-2000
pieces.
ACCESSION AC019884
VERSION AC019884.1 GI:6665013
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 54186)
AUTHORS Adams, M. and Venter, J. C.
TITLE Direct Submission
JOURNAL Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211271 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
Location/Qualifiers
1. 54186
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 15390 a 10935 c 11441 g 16420 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 2; Length 54186;
Best Local Similarity 94.7%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19
|||||
Db 12687 GCTGGAGGATTCCTCTATC 12669

RESULT 15
AC004564/c
LOCUS
DEFINITION Drosophila melanogaster DNA sequence (PI DS08012 (D224)), complete
sequence.
ACCESSION AC004564
VERSION AC004564.1 GI:3108013
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in Sac from PI clone DS08012
(D224)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 67833)

```

---

```

AUTHORS
Celniker, S.E., Aghavani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Humtani, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
Lomotan, M.A., Mada, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M.,
Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, D., Snir, E.,
Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R., Zieran, L.L.
and Kimmel, B.
Sequencing of Drosophila chromosome 2R, region 57E1-57E2
Unpublished (1997)
2 (bases 1 to 67833)
AUTHORS
Celniker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,
Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Humtani, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,
Zieran, L.L. and Kimmel, B.E.
Direct Submission
Submitted (17-APR-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On or before May 2, 1998 this sequence version replaced gi:2708065,
gi:2708064, gi:2708060, gi:2708059, gi:2708054, gi:3056946.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site
(http://fruitfly.berkeley.edu/sequence-archive.html) or send email
to drosophila@hgsc.lbl.gov.
Library location: 44-84.
FEATURES
source
Location/Qualifiers
1. 67833
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
/map="57E1-57E2"
/clone="PI DS08012 (D224)"
BASE COUNT 19352 a 13785 c 14011 g 20685 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 3; Length 67833;
Best Local Similarity 94.7%; Pred. No. 90;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19
|||||
Db 9820 GCTGGAGGATTCCTCTATC 9802

Search completed: July 8, 2003, 03:34:27
Job time : 225.098 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 125.659 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-7

Perfect score: 20

Sequence: 1 gctggaggattctctgtcc 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	N_Geneseq_101002.*
1:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	20	100.0	20	AAZ28813	Rat membrane metal
2	20	100.0	2765	AAZ28810	Rat membrane metal
3	17.4	87.0	2639	ABLI6565	Drosophila melanog
4	17.4	87.0	4716	ABLI6564	Drosophila melanog
5	17	85.0	566	AAH55840	Human SCN3A genom
6	17	85.0	6822	ABK63697	Rat sequence diffe
7	17	85.0	9112	AAH55823	Human adult form o
8	17	85.0	9112	AAH55824	Human neonatal for
9	17	85.0	9123	ABL39690	Human sodium chann

10	16.8	84.0	2286	24	AAZ28130	Soluble secreted e
11	16.8	84.0	2729	21	AAA38443	Murine desmin gene
12	16.8	84.0	2730	21	AAZ93780	Mouse desmin promo
13	16.4	82.0	1038602	20	AAZ01425	Complete genome se
14	16	80.0	245	21	AAC16982	Human secreted pro
15	16	80.0	1537	21	AAC39240	Arabidopsis thalia
16	16	80.0	2130	22	AAH44662	Equine infectious
17	16	80.0	2592	21	AAC66316	EIAV env gene SEQ
18	16	80.0	2598	22	AAH44655	Equine infectious
19	16	80.0	8236	22	AAH44652	Equine infectious
20	16	80.0	8255	22	AAH44659	Equine infectious
21	16	80.0	8258	21	AAC66281	Human SNP oligonuc
22	15.8	79.0	281	22	ABAL3433	Human nervous syst
23	15.8	79.0	303	23	ABL22193	Drosophila melanog
24	15.8	79.0	306	23	AAH48004	Enterococcus faeca
25	15.8	79.0	311	23	AAH47896	Enterococcus faeca
26	15.8	79.0	311	23	AAH48093	Enterococcus faeca
27	15.8	79.0	311	23	AAH48219	Human nervous syst
28	15.8	79.0	464	22	ABAL1480	Human brain expres
29	15.8	79.0	533	22	AAK10510	Human bone marrow
30	15.8	79.0	533	22	AAK36409	Enterococcus faeca
31	15.8	79.0	639	23	AAH51252	Aspergillus oryzae
32	15.8	79.0	670	21	AAF13754	Enterococcus faeca
33	15.8	79.0	696	23	AAH52802	Human nervous syst
34	15.8	79.0	760	22	ABAL19508	Insulin receptor g
35	15.8	79.0	928	22	AAH50547	Human insulin rece
36	15.8	79.0	928	22	AAH31147	Exon 2 of human in
37	15.8	79.0	928	22	AAH31244	Human insulin rece
38	15.8	79.0	928	22	AAH04444	Drosophila melanog
39	15.8	79.0	928	24	ABQ72702	Human secreted pro
40	15.8	79.0	1443	23	ABL21733	Hybrid human insul
41	15.8	79.0	1727	20	AAZ27243	Drosophila melanog
42	15.8	79.0	1800	12	AAQ14817	Cad-c fusion polype
43	15.8	79.0	2436	23	ABL22192	
44	15.8	79.0	2533	20	AAH58149	
45	15.8	79.0				

ALIGNMENTS

RESULT 1

ID	AAZ28813	standard; DNA; 20 BP.
XX	AAZ28813;	
AC	AAZ28813;	
DT	01-FEB-2000	(first entry)
XX		
DE	Rat membrane metalloprotease NEPII gene probe #3.	
XX		
KW	Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;	
KW	neuron; hormone; peptide messenger; inhibitor; detection; disorder;	
KW	cardiovascular disease; neurodegenerative disease; growth disorder;	
KW	hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.	
XX	Synthetic.	
OS	Rattus rattus.	
XX		
PN	FR2777291-Al.	
XX		
PD	15-OCT-1999.	
XX		
PF	08-APR-1998;	98FR-0004389.
XX		
PR	08-APR-1998;	98FR-0004389.
XX		
PA	(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.	
XX		
PI	Quimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;	
PI	Schwartz JC;	
XX		
DR	WPI; 1999-593429/51.	

XX New membrane metalloprotease NEP II, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 3; Page 20; 29pp; French.  
 XX  
 CC Sequences AA228811-228827 represent probes for detecting the rat  
 CC membrane metalloprotease designated neprilysine II (NEPII) gene  
 CC (AA228810). NEPII is involved in (in)activation of neuronal and hormonal  
 CC peptide messengers. NEPII is used to screen for specific substrates (used  
 CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
 CC used to detect NEPII or for treatment of disorders related to peptidergic  
 CC signalling in which NEPII is involved, e.g. cardiovascular or  
 CC neurodegenerative diseases; growth disorders of endocrine origin;  
 CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
 CC disorders.  
 XX  
 SQ Sequence 20 BP; 2 A; 6 C; 6 G; 6 T; 0 other;  
 Query Match 100.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 3.5;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 20  
 DB 1 GCTGGAGGATTCCTCTGTC 20  
 RESULT 2  
 AA228810  
 ID AA228810 standard; cDNA; 2765 BP.  
 XX  
 AC AA228810;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Rat membrane metalloprotease NEPII gene.  
 XX  
 KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
 KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
 KW cardiovascular disease; neurodegenerative disease; growth disorder;  
 KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
 XX  
 OS Rattus rattus.  
 XX  
 FR2777291-A1.  
 XX  
 PD 15-OCT-1999.  
 XX  
 PF 08-APR-1998; 98FR-0004389.  
 XX  
 PR 08-APR-1998; 98FR-0004389.  
 XX  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
 PI Schwartz JC;  
 XX  
 DR WPI: 1999-593429/51.  
 DR P-PSDB; AAY44177.  
 XX  
 PT New membrane metalloprotease NEP II, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 2; Page 12-16; 29pp; French.  
 XX  
 CC This sequence represents the gene for the rat membrane metalloprotease  
 CC designated neprilysine II (NEPII), which is involved in (in)activation  
 CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
 CC for specific substrates (used to detect NEPII in cells and tissues) or  
 CC inhibitors, which can also be used to detect NEPII or for treatment of

CC disorders related to peptidergic signalling in which NEPII is involved,  
 CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
 CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
 CC endocrine disorders.  
 XX  
 SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
 Query Match 100.0%; Score 20; DB 20; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 4.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 20  
 DB 553 GCTGGAGGATTCCTCTGTC 572  
 RESULT 3  
 ABL16565/C  
 ID ABL16565 standard; DNA; 2639 BP.  
 XX  
 AC ABL16565;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 1168.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions  
 XX  
 PS Claim 1; SEQ ID NO 1168; 21pp + Sequence Listing: English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2639 BP; 772 A; 614 C; 643 G; 610 T; 0 other;  
 Query Match 87.0%; Score 17.4; DB 23; Length 2639;  
 Best Local Similarity 94.7%; Pred. No. 81;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 19  
 DB 794 GCTGGAGGATTCCTCTATC 776

## RESULT 4

ABL1564  
ID ABL1564 standard; DNA; 4716 BP.  
XX  
AC ABL1564;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 1165.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
XX  
PS Claim 1; SEQ ID NO 1165; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABU01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 4716 BP; 1298 A; 1022 C; 1009 G; 1387 T; 0 other;

Query Match 87.0%; Score 17.4; DB 23; Length 4716;  
Best Local Similarity 94.7%; Pred. No. 84;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
|||||  
DB 2846 GCTGGAGGATTCCTCTATC 2864

## RESULT 5

AAH55840  
ID AAH55840 standard; DNA; 566 BP.  
XX  
AC AAH55840;  
XX  
DT 04-SEP-2001 (first entry)  
XX  
DE Human SCN3A genomic DNA fragment SEQ ID NO:84.  
XX  
KW Human; epilepsy; chromosome 2; SCN1A; SCN2A; SCN3A; identification;  
KW diagnosis; mutation; chromosome 2q23-q31; neurological disorder;

KW anticonvulsant; neuroprotective; ds.

XX Homo sapiens.

PN WO200138564-A2.

XX 31-MAY-2001.

XX 24-NOV-2000; 2000WO-CA01404.

XX 26-NOV-1999; 99US-0167623.

XX (UYMC-) UNIV MCGILL.

XX Rouleau GA, Lafreniere RG, Rochefort D, Cossette P, Ragsdale D;

XX WPI; 2001-355945/37.

XX Determining a predisposition to epilepsy and/or development of epilepsy  
PT comprises determining the genotype of SCN1A, SCN2A and/or SCN3A, or a  
PT DNA variant, equivalent, or mutation which shows a linkage  
PT disequilibrium

XX Disclosure; Page 179; 268pp; English.

XX The present invention describes a method (M1) of determining an  
CC individual's predisposition to epilepsy and/or development of epilepsy,  
CC as well as predicting the individual's response to medication. The  
CC method comprises determining the genotype of at least one gene selected  
CC from SCN1A, SCN2A or SCN3A, or a DNA variant, equivalent, or mutation  
CC which shows a linkage disequilibrium. SCN1A, SCN2A and SCN3A are all  
CC sodium channel genes located on chromosome 2. The idiopathic generalised  
CC epilepsy (IGE) gene is more specifically localised on chromosome  
CC 2q23-q31. Compounds identified as modulators of the biological activity  
CC of SCN1A, SCN2A or SCN3A proteins or genes, are useful for treating  
CC epilepsy or other neurological disorders. They have anticonvulsant and  
CC neuroprotective activities. AAH55763 to AAH56164 and AAH99674 to  
CC AAH99679 represent SCN1A, SCN2A, and SCN3A cDNAs, gene fragments, PCR  
CC primers, oligonucleotides and proteins given in the exemplification of  
CC the present invention.

XX Sequence 566 BP; 174 A; 93 C; 111 G; 187 T; 1 other;

Query Match 85.0%; Score 17; DB 22; Length 566;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTG 17  
|||||

DB 260 GCTGGAGGATTCCTCTG 276

## RESULT 6

ABK63697  
ID ABK63697 standard; cDNA; 6822 BP.

XX ABK63697;

XX 18-JUN-2002 (first entry)

XX Rat sequence differentially expressed in response to a hepatotoxin #1604.

XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
KW differential expression; centrilobular necrosis; steatosis.

XX Rattus norvegicus.

XX WO200210453-A2.

XX 07-FEB-2002.

XX 30-JUL-2001; 2001WO-US23872.

XX

PR 31-JUL-2000; 2000US-222040P.  
 PR 02-NOV-2000; 2000US-244880P.  
 PR 11-MAY-2001; 2001US-290029P.  
 PR 15-MAY-2001; 2001US-290645P.  
 PR 22-MAY-2001; 2001US-292336P.  
 PR 06-JUN-2001; 2001US-295798P.  
 PR 13-JUN-2001; 2001US-297457P.  
 PR 19-JUN-2001; 2001US-298884P.  
 PR 09-JUL-2001; 2001US-303459P.  
 XX (GENE-) GENE LOGIC INC.  
 PA  
 XX  
 XX  
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
 XX  
 XX  
 DR WPT; 2002-241625/29.  
 XX  
 XX

Predicting toxic effects of compounds or the progression of these toxic effects by determining the changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells

Claim 17: Seq ID No 1604; 239pp; English.

The invention relates to methods for predicting toxic effects of compounds or the progression of these toxic effects by determining the global changes in gene expression in tissues or cells exposed to the toxin and comparing these to gene expression in unexposed tissues or cells. Also included are methods of predicting at least one toxic effect of a compound or progression of a toxic effect, preferably the hepatotoxicity of a compound, comprising detecting the level of expression in a tissue or cell sample exposed to the compound of two or more genes listed in the specification, where differential expression of the genes is indicative of at least one toxic effect or progression. The method can also be used to identify an agent which modulates the toxic response and predict cellular pathways that a compound modulates in a cell. The methods utilise a set of at least two probes (on a solid support in kit form), where each of the probes comprises a sequence that specifically hybridises to a gene listed in the specification, a computer system comprising a database containing information identifying the expression level in a tissue or cell sample exposed to a hepatotoxin of a set of genes comprising at least two genes listed in the specification, and a user interface to view the information used to present information identifying the expression level in a tissue or cell of at least one gene listed in the specification. The method is useful for elucidating global changes in gene expression and for identifying toxicity markers in tissues or cell exposed to a known toxin. The genes may be used as toxicity markers in drug screening and toxicity assays. The genes and gene expression information may be used as diagnostic markers for the prediction or identification of the physiological state of tissue or cell sample that has been exposed to a compound or agent. Hepatotoxicity is characterised by centrilobular necrosis and steatosis. The present sequence is an expressed sequence tag (EST) or cDNA derived from a gene which is differentially expressed in response to a hepatotoxic agent.

Sequence 6822 BP; 1861 A; 1525 C; 1650 G; 1786 T; 0 other;

Query Match 85.0%; Score 17; DB 24; Length 6822;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTGGAGGATTCCTCTG 17  
 |||||  
 Db 2351 GCTGGAGGATTCCTCTG 2367

RESULT 7  
 AAH5823  
 ID AAH5823 standard; cDNA; 9112 BP.  
 XX  
 AC AAH5823;  
 XX  
 XX  
 DT 04-SEP-2001 (first entry)  
 XX

DE Human adult form of SCN3A nucleotide sequence SEQ ID NO:65.  
 XX  
 KW Human; epilepsy; chromosome 2; SCN1A; SCN2A; SCN3A; identification;  
 KW diagnosis; mutation; chromosome 2q23-q31; neurological disorder;  
 KW anticonvulsant; neuroprotective; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200138564-A2.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 24-NOV-2000; 2000WO-CA01404.  
 XX  
 XX 26-NOV-1999; 99US-0167623.  
 PR (UYMC-) UNIV MCGILL.  
 XX  
 XX Rouleau GA, LaFreniere RG, Rochefort D, Cossette P, Ragsdale D;  
 PI  
 XX WPI; 2001-355945/37.  
 DR P-PSDB; AAB99678.  
 XX

Determining a predisposition to epilepsy and/or development of epilepsy comprises determining the genotype of SCN1A, SCN2A and/or SCN3A, or a DNA variant, equivalent, or mutation which shows a linkage disequilibrium

Disclosure; Page 150-154; 268pp; English.

The present invention describes a method (M1) of determining an individual's predisposition to epilepsy and/or development of epilepsy, as well as predicting the individual's response to medication. The method comprises determining the genotype of at least one gene selected from SCN1A, SCN2A or SCN3A, or a DNA variant, equivalent, or mutation which shows a linkage disequilibrium. SCN1a, SCN2A and SCN3A are all sodium channel genes located on chromosome 2. The idiopathic generalised epilepsy (IGE) gene is more specifically localised on chromosome 2. The idiopathic generalised epilepsy or other neurological disorders. They have anticonvulsant and neuroprotective activities. AAH55763 to AAH56164 and AAB99674 to AAB99679 represent SCN1A, SCN2A, and SCN3A cDNAs, gene fragments, PCR primers, oligonucleotides and proteins given in the exemplification of the present invention.

Sequence 9112 BP; 2769 A; 1695 C; 1861 G; 2785 T; 2 other;

Query Match 85.0%; Score 17; DB 22; Length 9112;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTGGAGGATTCCTCTG 17  
 |||||  
 Db 2570 GCTGGAGGATTCCTCTG 2586

RESULT 8  
 AAH5824  
 ID AAH5824 standard; cDNA; 9112 BP.  
 XX  
 AC AAH5824;  
 XX

DT 04-SEP-2001 (first entry)  
 XX  
 XX

DE Human neonatal form of SCN3A nucleotide sequence SEQ ID NO:66.  
 XX  
 KW Human; epilepsy; chromosome 2; SCN1A; SCN2A; SCN3A; identification;  
 KW diagnosis; mutation; chromosome 2q23-q31; neurological disorder;  
 KW anticonvulsant; neuroprotective; ss.  
 XX  
 OS Homo sapiens.  
 XX

PN WO200138564-A2.  
XX 31-MAY-2001.  
PD 24-NOV-2000; 2000WO-CA01404.  
XX 26-NOV-1999; 99US-0167623.  
XX (UYMC-) UNIV MCGILL.  
XX Rouleau GA, LaFreniere RG, Rochefort D, Cossette P, Ragsdale D;  
PI WPI; 2001-355945/37.  
XX P-PSDB; AAB99679.  
DR Determining a predisposition to epilepsy and/or development of epilepsy  
PT comprises determining the genotype of SCN1A, SCN2A and/or SCN3A, or a  
PT DNA variant, equivalent, or mutation which shows a linkage  
PT disequilibrium.  
XX Disclosure; Page 154-157; 268pp; English.  
PS The present invention describes a method (M1) of determining an  
CC individual's predisposition to epilepsy and/or development of epilepsy,  
CC as well as predicting the individual's response to medication. The  
CC method comprises determining the genotype of at least one gene selected  
CC from SCN1A, SCN2A or SCN3A, or a DNA variant, equivalent, or mutation  
CC which shows a linkage disequilibrium. SCN1A, SCN2A and SCN3A are all  
CC sodium channel genes located on chromosome 2. The idiopathic generalised  
CC epilepsy (IGE) gene is more specifically localised on chromosome  
CC 2q23-q31. Compounds identified as modulators of the biological activity  
CC of SCN1A, SCN2A or SCN3A proteins or genes, are useful for treating  
CC epileptic or other neurological disorders. They have anticonvulsant and  
CC neuroprotective activities. AAH55763 to AAH56164 and AAB99674 to  
CC AAB99679 represent SCN1A, SCN2A, and SCN3A cDNAs, gene fragments, PCR  
CC primers, oligonucleotides and proteins given in the exemplification of  
CC the present invention.  
XX Sequence 9112 BP; 2769 A; 1695 C; 1857 G; 2789 T; 2 other;  
SQ Query Match 85.0%; Score 17; DB 22; Length 9112;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCTGGAGGATTCCTCTG 17  
Db 2570 GCTGGAGGATTCCTCTG 2586  
RESULT 9  
ABL39690  
ID ABL39690 standard; cDNA; 9123 BP.  
XX ABL39690;  
AC ABL39690;  
XX 10-MAY-2002 (first entry)  
DT Human sodium channel SCN3A encoding cDNA SEQ ID NO:3.  
DE Human; sodium channel; SCN3A; chromosome 2q24-31; gene; ss;  
KW familial hypercalcaemic periodic paralysis; motor endplate disease.  
KW Homo sapiens.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH 472..6474  
FT /\*tag= a  
FT /product= "sodium channel SCN3A"  
XX WO200196552-A1.  
PN 20-DEC-2001.  
XX

PF 12-JUN-2001; 2001WO-JP04956.  
XX 13-JUN-2000; 2000JP-0177540.  
PR 13-JUN-2000; 2000JP-0177544.  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX Kanazawa I; Goto J, Jeong S;  
PI WPI; 2002-098066/13.  
XX P-PSDB; ABB06027.  
DR Human sodium channels SCN1A and SCN3A and encoded genes, useful in  
XX studying physiological mechanism in which excitant cells participate  
XX and causes of diseases and developing drugs for motor endplate disease  
PT Claim 6; Page 50-71; 88pp; Japanese.  
PS The present invention describes human sodium channels SCN1A and SCN3A.  
XX The present sequence encodes the human sodium channel SCN3A. SCN1A and  
CC SCN3A have been located to the human chromosome 2 long arm, positions  
CC 2q24 and 2q24-31 respectively. The sodium channel proteins are useful  
CC in studying the physiological mechanism in which excitant cells  
CC participate and cause human diseases, and in developing remedies for  
CC e.g. familial hypercalcaemic periodic paralysis of extremities and  
CC motor endplate disease.  
XX Sequence 9123 BP; 2770 A; 1700 C; 1870 G; 2783 T; 0 other;  
SQ Query Match 85.0%; Score 17; DB 24; Length 9123;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCTGGAGGATTCCTCTG 17  
Db 2556 GCTGGAGGATTCCTCTG 2572  
RESULT 10  
AAD28130  
ID AAD28130 standard; DNA; 2286 BP.  
XX AAD28130;  
AC AAD28130;  
XX 07-MAY-2002 (first entry)  
DT Soluble secreted endopeptidase (SEP) consensus DNA.  
DE Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
XX enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
KW PSAD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
XX Homo sapiens.  
OS Mus sp.  
OS Rattus sp.  
XX Key Location/Qualifiers  
FH misc\_feature 1664..2286  
FT /\*tag= a  
FT /note= "Encodes catalytic domain"  
XX WO200206492-A1.  
PN 24-JAN-2002.  
XX 16-JUL-2001; 2001WO-IB01263.  
PF 14-JUL-2000; 2000GB-0017387.  
XX (PF12 ) PFIZER LTD.  
XX (PF12 ) PFIZER INC.  
PA

98EP-0402435.

```

1 FT
2 FT
3 FT
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
10
```

XX (UYPA-) UNIV PARIS VII.  
 XX Paulin D, Li Z, Mericskay M;  
 XX WPI; 2000-306012/27.  
 XX Novel nucleic acids comprising the 5' flanking region of a desmin gene  
 PT located useful producing drugs and targeting gene expression -  
 FT  
 PS Claim 6; Fig 1; 23pp; English.  
 XX  
 CC This sequence represents the murine desmin gene promoter, comprising the  
 CC region between nucleotides -3168 to +83 relative to the transcription to  
 CC initiation site. This sequence contains regulatory elements specific to  
 CC smooth, cardiac and skeletal muscle cells, and includes CARG-like  
 CC sequences. CARG motifs are elements known to regulate many muscle-  
 CC specific genes, especially the alpha-actin gene in cardiac and  
 CC skeletal muscle. Sequence comparison between the murine and  
 CC human desmin gene 5' flanking sequences (AAA38443, AAA38444) shows four  
 CC conserved regions between the two species, one of which is located in  
 CC the arterial smooth muscle-specific region. Murine desmin 5' flanking  
 CC region sequences may be used to construct expression cassettes for  
 CC tissue-specific expression of a gene of interest. The sequence of at  
 CC least part of -10000 to +600 (preferably -4006 to +603) of the mouse  
 CC desmin gene (or a homologous region of the human desmin gene) is used to  
 CC target the expression of a gene to cardiac, skeletal, or smooth muscle,  
 CC and at least part of -5000 to -2400 (preferably -4006 to -2495  
 CC (AAA38442)) of the mouse desmin gene is used to target the expression of  
 CC a gene to a minimal promoter in an arterial smooth muscle. The expression  
 CC cassettes, and vectors and eukaryotic cells comprising the expression of  
 CC cassettes are useful in gene therapy for the treatment or prevention of  
 CC cardiovascular diseases. The cardiovascular diseases include restenosis,  
 CC ischaemia, intimal hyperplasia, hypercholesterolaemia, atherosclerosis  
 CC and smooth muscle proliferation.  
 XX  
 SQ Sequence 2729 BP; 639 A; 708 C; 720 G; 659 T; 3 other;  
 Query Match 84.0%; Score 16.8; DB 21; Length 2729;  
 Best Local Similarity 90.0%; Pred.No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GCTGAGGATTCCTCTGCC 20  
 Db 2571 GCTGAGGATTCCTCTGCC 2552  
 RESULT 12  
 AA293780/C  
 ID AA293780 standard; DNA; 2730 BP.  
 AC AA293780;  
 XX  
 XX 16-AUG-2000 (first entry)  
 XX  
 DE Mouse desmin promoter sequence.  
 XX  
 KW Desmin gene; 5' flanking region; promoter; cardiovascular disease;  
 KW smooth muscle cells; restenosis; ischemia; intimal hyperplasia;  
 KW atherosclerosis; hypercholesterolemia;  
 KW smooth muscle cell proliferation; cancer; viral hepatitis;  
 KW antisense; ribozyme; transgenic animal; gene expression; artery;  
 KW aorta; pulmonary artery; mouse; ss.  
 XX  
 OS Mus musculus.  
 XX  
 XX Key Location/Qualifiers  
 FH misc\_binding 223..521  
 FT /tag= a  
 FT /note= "Putative consensus binding site for  
 FT transcription factor"  
 FT misc\_binding 338..340  
 FT /tag= b

FT misc\_binding  
 FT /label= Consensus binding site for NF1  
 FT 347..351  
 FT /tag= c  
 FT /label= Consensus binding site for NF1  
 FT 373..380  
 FT /tag= d  
 FT /label= Consensus binding site for AP-1  
 FT 397..403  
 FT /tag= e  
 FT /label= Consensus binding site for HNF-5  
 FT 597..606  
 FT /tag= f  
 FT /label= CARG-2 box motif  
 FT 973..1209  
 FT /tag= g  
 FT /note= "Putative consensus binding site for  
 FT transcription factor"  
 FT misc\_feature 1210  
 FT /tag= h  
 FT /note= "N corresponds to a 522 base pair sequence  
 FT not reproduced in specification"  
 FT misc\_feature 1258..1267  
 FT /tag= i  
 FT /label= CARG-1  
 FT 1435..1442  
 FT /tag= j  
 FT /label= C/EBP  
 FT 1451..1458  
 FT /tag= k  
 FT /label= Bicoid  
 FT 1693..1699  
 FT /tag= l  
 FT /label= PEA3  
 FT 1722..2001  
 FT /tag= m  
 FT /note= "Putative consensus binding site for  
 FT transcription factor"  
 FT misc\_binding 1789..1799  
 FT /tag= n  
 FT /label= MEF-2 Binding site  
 FT 1793..1799  
 FT /tag= o  
 FT /label= Myotube enhancer Mt site  
 FT 1806..1812  
 FT /tag= p  
 FT /label= Myotube enhancer Mt site  
 FT 1821..1826  
 FT /tag= q  
 FT /label= MyoD1  
 FT 1889..1898  
 FT /tag= r  
 FT /label= Krox 20/24 binding site  
 FT 1925..1933  
 FT /tag= s  
 FT /label= Krox 20/24 binding site  
 FT 1944..1955  
 FT /tag= t  
 FT /label= Mb  
 FT 1970..1978  
 FT /tag= u  
 FT /label= Krox 20/24  
 FT 2475..2729  
 FT /tag= v  
 FT /note= "Putative consensus binding site for  
 FT transcription factor"  
 FT misc\_feature 2569..2574  
 FT /tag= w  
 FT /label= MyoD1  
 FT 2612..2618  
 FT /tag= x  
 XX  
 PN AU9887889-A.  
 XX

PD 06-APR-2000.  
 XX  
 PF 02-OCT-1998; 98AU-0087889.  
 XX  
 PR 02-OCT-1998; 98AU-0087889.  
 XX  
 PA (UYPA-) UNIV PARIS 7.  
 XX  
 PI Paulin D, Li Z, Mericskay M;  
 XX WPI; 2000-318229/28.  
 DR  
 XX New regulatory region from the desmin gene, useful in gene therapy,  
 PT particularly of cardiovascular disease, provides expression  
 PT specifically in arterial smooth muscle cells  
 XX  
 PS Claim 1; Figure 1; 37pp; English.  
 XX  
 CC The desmin gene encodes a cytoskeletal protein constitutive of  
 CC intermediate filaments which occur in the cytoplasm of most muscular  
 CC cells. Studies have revealed a complex regulatory mechanism to  
 CC promote the temporal and tissue-specific expression of the desmin  
 CC gene during myogenesis. Positive and negative control sequences have  
 CC been described in the first 1000 bp of the 5' flanking region of the  
 CC desmin gene. It is thought therefore that the desmin gene promoter  
 CC region can be used to target expression in vascular smooth muscle  
 CC cells and this might be beneficial to avoid the potential side  
 CC effects inherent to a widespread gene expression. The desmin gene 5'  
 CC flanking region can be used to prepare expression cassettes  
 CC containing a selected gene for the controllable expression of that  
 CC gene. Such methods could be used for gene therapy, of cardiovascular  
 CC disease (specifically restenosis, ischemia, intimal hyperplasia,  
 CC atherosclerosis, hypercholesterolemia and smooth muscle cell  
 CC proliferation), for expression of a therapeutic polypeptide,  
 CC antisense molecule or ribozyme, e.g. in cases of genetic disorders,  
 CC cancer, viral hepatitis etc. The expression cassettes could also be  
 CC used to produce transgenic animals, useful as models for studying  
 CC effects of the desmin gene 5' flanking region and heterologous genes  
 CC to be expressed in such expression cassettes. The advantage of using  
 CC the desmin gene 5' flanking region is that it targets expression of  
 CC attached genes to cardiac, skeletal and smooth muscle cells,  
 CC especially arterial smooth muscle cells and specifically those in the  
 CC aorta and pulmonary artery.  
 XX  
 SQ Sequence 2730 BP; 639 A; 715 C; 714 G; 661 T; 1 other;  
 Query Match 84.0%; Score 16.8; DB 21; Length 2730;  
 Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GCTGAGGAGTCCCTCTGCC 20  
 DB 2572 GCTGAGGAGTCCCTCTGCC 2553  
 RESULT 13  
 AAZ01425  
 ID AAZ01425 standard; DNA; 1038602 BP.  
 XX  
 AC AAZ01425;  
 XX  
 DT 07-OCT-1999 (first entry)  
 XX  
 DE Complete genome sequence of Chlamydia trachomatis.  
 XX  
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO9928475-A2.

XX 10-JUN-1999.  
 PD  
 XX 27-NOV-1998; 98WO-IB01939.  
 PF  
 XX 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 PA (GSET ) GENSET.  
 XX  
 PI Griffiths R;  
 XX  
 DR WPI; 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 PS  
 XX Claim 1; Page 373-656; 1755pp; English.  
 CC  
 CC The present sequence represents the complete genome of Chlamydia  
 CC trachomatis. Open reading frames (ORFs) of the genome encode  
 CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines  
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also  
 CC be used to control growth of the microorganism. Chlamydia trachomatis is  
 CC responsible for a large number of diseases, e.g. eye diseases such as  
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion  
 CC conjunctivitis; genital diseases such as nongonococcal urethritis,  
 CC epididymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis;  
 CC pneumopathy in breast feeding infants; and venereal  
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in  
 CC treating these diseases.  
 XX  
 SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;  
 Query Match 82.0%; Score 16.4; DB 20; Length 1038602;  
 Best Local Similarity 94.4%; Pred. No. 3.2e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CTGAGGAGTCCCTCTGTC 19  
 DB 450775 CAGGAGGATCCCTCTGTC 450792  
 RESULT 14  
 AAC16982/c  
 ID AAC16982 standard; cDNA; 245 BP.  
 XX  
 AC AAC16982;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 21057.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GSET ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX WPI; 2000-500381/45.  
 DR  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 21057; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.

XX SQ Sequence 245 BP; 61 A; 67 C; 52 G; 65 T; 0 other;

Query Match 80.0%; Score 16; DB 21; Length 245;

Best Local Similarity 100.0%; Pred. No. 3.3e+02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 GGAGGATTCCTCTGTC 19

|||||

Db 105 GGAGGATTCCTCTGTC 90

#### RESULT 15

AAC39240

ID AAC39240 standard; DNA; 1537 BP.

XX AAC39240;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23888.

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140895.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154773.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159233.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 21-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.

PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 80.0%; Score 16; DB 21; Length 1537;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TGGAGGATTCTCTGT 18  
|||||  
Db 1153 TGGAGGATTCTCTGT 1168

Search completed: July 8, 2003, 02:18:53  
Job time : 128.659 secs



STREET: 2100 Pennsylvania Avenue, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/737,715  
FILING DATE: 12-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 134827/1995  
FILING DATE: 12-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Nakamura, Dean H.  
REGISTRATION NUMBER: 33,981  
REFERENCE/DOCKET NUMBER: Q-43323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)293-7060  
TELEFAX: (202)293-7860  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4149 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..4146  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 82..4146  
US-08-737-715-1

Query Match 79.0%; Score 15.8; DB 2; Length 4149;  
Best Local Similarity 89.5%; Pred. No. 67;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||||| ||| |||||  
Db 147 GCTGGAGGATTCCTCTGTC 165

RESULT 3  
US-09-301-665-3  
; Sequence 3, Application US/09301665  
; Patent No. 6207876  
; GENERAL INFORMATION:  
; APPLICANT: KELLEMS, RODNEY E.  
; APPLICANT: DATA, SURJIT K.  
; APPLICANT: BLACKBURN, MICHAEL R.  
; TITLE OF INVENTION: ADENOSINE DEAMINASE DEFICIENT TRANSGENIC MICE AND  
; FILE REFERENCE: METHODS FOR THE USE THEREOF  
; CURRENT APPLICATION NUMBER: US/09/301,665  
; CURRENT FILING DATE: 1999-04-28  
; EARLIER APPLICATION NUMBER: 60/083,408  
; EARLIER FILING DATE: 1998-04-29  
; EARLIER APPLICATION NUMBER: 60/083,370  
; EARLIER FILING DATE: 1998-04-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 36741  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-301-665-3

Query Match 79.0%; Score 15.8; DB 4; Length 36741;  
Best Local Similarity 89.5%; Pred. No. 80;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||||| ||| |||||  
Db 10280 GCTGGAGGATTCCTCTGTC 10298

RESULT 4  
US-09-800-960-3  
; Sequence 3, Application US/09800960  
; Patent No. 6387677  
; GENERAL INFORMATION:  
; APPLICANT: YE, Jane et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL001158  
; CURRENT APPLICATION NUMBER: US/09/800,960  
; CURRENT FILING DATE: 2001-03-08  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 62804  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(62804)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-800-960-3

Query Match 79.0%; Score 15.8; DB 4; Length 62804;  
Best Local Similarity 89.5%; Pred. No. 83;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||||| ||| |||||  
Db 56219 GCTGGAGGATTCCTCTGTC 56237

RESULT 5  
US-09-574-942-1  
; Sequence 1, Application US/09574942  
; Patent No. 6358723  
; GENERAL INFORMATION:  
; APPLICANT: Eastman, Alan R.  
; APPLICANT: Krieser, Ronald J.  
; TITLE OF INVENTION: DEOXYRIBONUCLEASE II BETA PROTEINS AND CDNAS  
; FILE REFERENCE: DC-0137.  
; CURRENT APPLICATION NUMBER: US/09/574,942  
; CURRENT FILING DATE: 2000-05-19  
; PRIOR APPLICATION NUMBER: 09/541,840  
; PRIOR FILING DATE: 2000-04-03  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1224  
; TYPE: DNA  
; ORGANISM: Mus. sp.  
US-09-574-942-1

Query Match 76.0%; Score 15.2; DB 4; Length 1224;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 20  
||||||| ||| |||||  
Db 1021 GGTGGAGGATTCATCTGTAC 1040

RESULT 6

US-08-339-152A-15/c  
; Sequence 15, Application US/08339152A  
; Patent No. 5643726  
; GENERAL INFORMATION:  
; APPLICANT: Tanzi, Rudolph E.  
; APPLICANT: Kovacs, Dora M.  
; TITLE OF INVENTION: Methods For Modulating Transcription  
; TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
; STREET: 1100 New York Ave., NW, Suite 600  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/339,152A  
; FILING DATE: 10-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Steffe, Eric K.  
; REGISTRATION NUMBER: 36,688  
; REFERENCE/DOCKET NUMBER: 0609.4120000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-2600  
; TELEFAX: 202-371-2540  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2358 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 88...2046  
; US-08-339-152A-15

Query Match 76.0%; Score 15.2; DB 1; Length 2358;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCTGGAGGATTCCTCTGTCC 20  
||||||| ||| |||||  
Db 1630 GCTGGAGGCTACCTTTGTCC 1611

RESULT 7  
US-08-007-999B-2/c  
; Sequence 2, Application US/08007999B  
; Patent No. 5851787  
; GENERAL INFORMATION:  
; APPLICANT: Wasco, Wilma  
; APPLICANT: Bupp, Keith  
; APPLICANT: Magendantz, Margaret  
; APPLICANT: Tanzi, Rudolph  
; APPLICANT: Solomon, Frank  
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.  
; STREET: 1100 New York Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA

ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/007,999B  
FILING DATE: 21-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,642  
FILING DATE: 20-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/930,022  
FILING DATE: 17-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Townsend, G. Kevin  
REGISTRATION NUMBER: 34,033  
REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)371-2571  
TELEFAX: (202)371-2540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2358 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 88...2046  
US-08-007-999B-2

Query Match 76.0%; Score 15.2; DB 2; Length 2358;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GCTGGAGGATTCCTCTGTCC 20  
||||||| ||| |||||  
Db 1630 GCTGGAGGCTACCTTTGTCC 1611

RESULT 8  
US-08-689-276A-2/c  
; Sequence 2, Application US/08689276A  
; Patent No. 5891991  
; GENERAL INFORMATION:  
; APPLICANT: Wasco, Wilma  
; APPLICANT: Bupp, Keith  
; APPLICANT: Magendantz, Margaret  
; APPLICANT: Tanzi, Rudolph  
; APPLICANT: Solomon, Frank  
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Ave., NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/689,276A  
; FILING DATE: 06-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/007,999

FILING DATE: 21-JAN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,642  
FILING DATE: 20-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/930,022  
FILING DATE: 17-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: JORGE A. GOLDSTEIN  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 0609.3520003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)371-2571  
TELEFAX: (202)371-2540  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2358 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 88..2046  
US-08-689-276A-2

Query Match 76.0%; Score 15.2; DB 2; Length 2358;  
Best Local Similarity 85.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GCTGGAGGATTCCTCTGCC 20  
Db 1630 GCTGGAGGCTACCTTTGTC 1611

RESULT 9  
US-09-754-250-3  
Sequence 3, Application US/09754250  
Patent No. 6376225  
GENERAL INFORMATION:  
APPLICANT: WEI, Ming-Hui et al  
TITLE OF INVENTION: ISOLATED HUMAN PHOSPHODIESTERASE  
TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
TITLE OF INVENTION: PHOSPHODIESTERASE PROTEINS, AND USES THEREOF  
FILE REFERENCE: CL001063  
CURRENT APPLICATION NUMBER: US/09/754,250  
CURRENT FILING DATE: 2001-01-05  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 111282  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(111282)  
OTHER INFORMATION: n - A,T,C or G  
US-09-754-250-3

Query Match 76.0%; Score 15.2; DB 4; Length 111282;  
Best Local Similarity 85.0%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GCTGGAGGATTCCTCTGCC 20  
Db 32099 GCTGGAGGATTCCTCTGCC 32118

RESULT 10  
US-08-609-443B-17/c  
Sequence 17, Application US/08609443B  
Patent No. 5840693  
GENERAL INFORMATION:

APPLICANT: ERIKSSON, Ulf  
APPLICANT: OLOFSSON, Birgitta  
APPLICANT: ALITALO, Kari  
APPLICANT: PAJUSOLA, Katri  
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,443B  
FILING DATE: 01-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/397,651  
FILING DATE: 01-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/469,427  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/569,063  
FILING DATE: 06-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/41979CP4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-609-443B-17

Query Match 74.0%; Score 14.8; DB 2; Length 1550;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GCTGGAGGATTCCTCTCT 18  
Db 920 GCTGAAGGATTCCTCTCT 903

RESULT 11  
US-08-851-896-17/c  
Sequence 17, Application US/08851896  
Patent No. 6331301  
GENERAL INFORMATION:  
APPLICANT: ERIKSSON, Ulf  
APPLICANT: OLOFSSON, Birgitta  
APPLICANT: ALITALO, Kari  
APPLICANT: PAJUSOLA, Katri  
TITLE OF INVENTION: VASCULAR ENDOTHELIAL GROWTH FACTOR-B AND  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan, P.L.L.C.  
STREET: 1200 G Street, N.W., Suite 700  
CITY: Washington

STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,896  
FILING DATE: 06-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/609,443B  
FILING DATE: 01-MAR-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/397,651  
FILING DATE: 01-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,427  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/569,063  
FILING DATE: 06-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: EVANS, Joseph D.  
REGISTRATION NUMBER: 26,269  
REFERENCE/DOCKET NUMBER: 1064/41979CP4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-8800  
TELEFAX: (202) 628-8844  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1550 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-851-896-17

Query Match 74.0%; Score 14.8; DB 4; Length 1550;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGT 18  
|||||  
DB 920 GCTGAAGGATGCTCTGT 903

RESULT 12  
US-08-303-569B-6/c  
Sequence 6, Application US/08303569B  
Patent No. 5859205  
GENERAL INFORMATION:  
APPLICANT: Adair, John R.  
APPLICANT: Athwal, Diljeet S.  
APPLICANT: Entage, John S.  
TITLE OF INVENTION: Humanised Antibodies  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5859205ris  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/303,569B  
FILING DATE: 07-SEP-1994

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Trujillo, Doreen Yanko  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: CARP-0032  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1570 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 41..1444  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 98..1444  
US-08-303-569B-6

Query Match 74.0%; Score 14.8; DB 2; Length 1570;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGT 18  
|||||  
DB 328 GCTGGAGGATTCCTCTGT 311

RESULT 13  
US-08-116-247-6/c  
Sequence 6, Application US/08116247  
Patent No. 5929212  
GENERAL INFORMATION:  
APPLICANT: Jolliffe, Linda K.  
APPLICANT: Zivvin, Robert A.  
APPLICANT: Adair, John R.  
APPLICANT: Athwal, Diljeet S.  
TITLE OF INVENTION: CD3 Specific Recombinant Antibody  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5929212ris  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/116,247  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/743,377  
FILING DATE: 10-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Paintin, Francis A.  
REGISTRATION NUMBER: 19,386  
REFERENCE/DOCKET NUMBER: CARP-0011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1570 base pairs  
TYPE: nucleic acid

;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 41..1444  
US-08-116-247-6

Query Match 74.0%; Score 14.8; DB 2; Length 1570;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGT 18  
|||||  
Db 328 GCTGGAGGATTCCTCTGT 311

RESULT 14  
US-09-091-725-14  
; Sequence 14, Application US/09091725  
; Patent No. 6329141  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Improved methods for transforming Phaffia  
; and recombinant DNA for use therein  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster llp  
; STREET: 2000 Pennsylvania Avenue, N.W.  
; CITY: Washington  
; STATE: DC  
; COUNTRY: United States of America  
; ZIP: 20006-1888  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/091,725  
; FILING DATE: 23-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 95203620.0  
; FILING DATE: 22-DEC-1995  
; APPLICATION NUMBER: EP 96200943.7  
; FILING DATE: 11-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: E. Victor Donahue  
; REGISTRATION NUMBER: 35,492  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1882 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Phaffia rhodozyma  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 82..1212  
; OTHER INFORMATION: /product= "Prate"  
US-09-091-725-14

Query Match 74.0%; Score 14.8; DB 4; Length 1882;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TGGAGGATTCCTCTGTCC 20  
|||||  
Db 308 TGGAGGATTCATCGTCC 325

RESULT 15  
US-08-272-255-5/c  
; Sequence 5, Application US/08272255  
; Patent No. 5824859  
; GENERAL INFORMATION:  
; APPLICANT: Cashmore, Anthony R.  
; APPLICANT: Ahmad, Margaret  
; APPLICANT: Lin, Chentao  
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
; TITLE OF INVENTION: Using the Same  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/272,255  
; FILING DATE: 08-JUL-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leary Ph.D., Kathryn  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: UPN-1795  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2458 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 148..2190  
US-08-272-255-5

Query Match 74.0%; Score 14.8; DB 1; Length 2458;  
Best Local Similarity 88.9%; Pred. No. 1.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGT 18  
|||||  
Db 2025-GCTGGAGATTCGCTGT 2008

Search completed: July 8, 2003, 09:31:02  
Job time : 29.6598 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 115.512 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-7  
Perfect score: 20  
Sequence: 1 gctggaggattcctctgtcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : Published Applications NA:\*
- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
  - 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
  - 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
  - 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
  - 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
  - 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
  - 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
  - 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.4	87.0	1323	9 US-10-108-605-46	Sequence 46, Appl
C 2	17	85.0	6822	10 US-09-917-800A-1604	Sequence 1604, Ap
C 3	15.8	79.0	306	10 US-09-815-242-581	Sequence 581, App
C 4	15.8	79.0	311	10 US-09-815-242-473	Sequence 473, App
C 5	15.8	79.0	311	10 US-09-815-242-670	Sequence 670, App
C 6	15.8	79.0	311	10 US-09-815-242-796	Sequence 796, App
C 7	15.8	79.0	533	10 US-09-864-761-13156	Sequence 13156, A
C 8	15.8	79.0	639	10 US-09-815-242-3834	Sequence 3834, Ap
C 9	15.8	79.0	696	10 US-09-815-242-6439	Sequence 6439, Ap
C 10	15.8	79.0	1727	9 US-09-798-889-21	Sequence 21, Appl
C 11	15.8	79.0	1901	10 US-09-874-132-22	Sequence 22, Appl
C 12	15.8	79.0	5180	10 US-09-954-456-1246	Sequence 1246, Ap
C 13	15.8	79.0	5180	10 US-09-880-107-3668	Sequence 3668, Ap
C 14	15.8	79.0	14557	10 US-09-070-927A-367	Sequence 367, App
C 15	15.8	79.0	36741	10 US-09-782-378A-12	Sequence 12, Appl
C 16	15.8	79.0	62804	12 US-10-096-960-3	Sequence 3, Appl
C 17	15.8	79.0	254366	9 US-09-822-871-3	Sequence 3, Appl
C 18	15.8	79.0	659158	9 US-09-771-208-20	Sequence 20, Appl
C 19	15.4	77.0	409	9 US-10-138-505-7	Sequence 7, Appl

C 20	15.4	77.0	819	9 US-10-138-505-29	Sequence 29, Appl
C 21	15.4	77.0	825	9 US-10-207-655-8	Sequence 8, Appl
C 22	15.4	77.0	828	9 US-10-138-505-25	Sequence 25, Appl
C 23	15.4	77.0	2473	9 US-10-198-846-12621	Sequence 12621, A
C 24	15.4	77.0	6633	12 US-10-044-090-531	Sequence 531, App
C 25	15.2	76.0	428	10 US-09-960-352-7795	Sequence 7795, Ap
C 26	15.2	76.0	429	10 US-09-983-965-5313	Sequence 5313, Ap
C 27	15.2	76.0	550	10 US-09-974-300-2561	Sequence 2561, Ap
C 28	15.2	76.0	721	9 US-10-023-282-141	Sequence 141, App
C 29	15.2	76.0	789	9 US-10-156-761-1164	Sequence 1164, Ap
C 30	15.2	76.0	1224	10 US-09-949-434-1	Sequence 1, Appl
C 31	15.2	76.0	1952	9 US-09-991-053-23	Sequence 23, Appl
C 32	15.2	76.0	1952	9 US-09-957-187-23	Sequence 23, Appl
C 33	15.2	76.0	2010	9 US-09-991-053-21	Sequence 21, Appl
C 34	15.2	76.0	2010	9 US-09-957-187-21	Sequence 21, Appl
C 35	15.2	76.0	2226	10 US-09-917-800A-1448	Sequence 1448, Ap
C 36	15.2	76.0	2496	9 US-09-808-602-75	Sequence 75, Appl
C 37	15.2	76.0	2496	9 US-09-800-198-63	Sequence 63, Appl
C 38	15.2	76.0	2589	9 US-09-938-842A-659	Sequence 659, App
C 39	15.2	76.0	3363	8 US-08-915-048A-1	Sequence 1, Appl
C 40	15.2	76.0	3317	10 US-09-925-302-182	Sequence 182, App
C 41	15.2	76.0	3692	9 US-09-798-889-15	Sequence 15, Appl
C 42	15.2	76.0	8409	9 US-09-808-602-79	Sequence 79, Appl
C 43	15.2	76.0	8409	9 US-09-800-198-67	Sequence 67, Appl
C 44	15.2	76.0	111282	12 US-10-094-989-3	Sequence 3, Appl
C 45	15.2	76.0	197997	10 US-09-822-246-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1  
US-10-108-605-46/c  
; Sequence 46, Application US/10108605  
; Patent No. US20020160934A1  
; GENERAL INFORMATION:  
; APPLICANT: Broadus, Julie  
; APPLICANT: Stam, Lynn  
; APPLICANT: Bachmann, Jane  
; APPLICANT: Kamdar, Kim  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCO  
; FILE REFERENCE: 31133B  
; CURRENT APPLICATION NUMBER: US/10/108,605  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 09/761,142  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/176,418  
; PRIOR FILING DATE: 2000-01-14  
; NUMBER OF SEQ ID NOS: 361  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 46  
; LENGTH: 1323  
; TYPE: DNA  
; ORGANISM: Drosophila melanogaster  
US-10-108-605-46

Query Match 87.0%; Score 17.4; DB 9; Length 1323;  
Best Local Similarity 94.7%; Pred. No. 18;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
|||||  
Db 686 GCTGGAGGATTCCTCTATC 668

RESULT 2  
US-09-917-800A-1604  
; Sequence 1604, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark

APPLICANT: Johnson, Kory  
APPLICANT: Castle, Arthur  
APPLICANT: Elashoff, Michael  
APPLICANT: Gene Logic, Inc.  
TITLE OF INVENTION: Molecular Toxicology Modeling  
FILE REFERENCE: 44921-5038-US  
CURRENT FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: US 60/222,040  
PRIOR FILING DATE: 2000-07-31  
PRIOR APPLICATION NUMBER: US 60/222,880  
PRIOR FILING DATE: 2000-11-02  
PRIOR APPLICATION NUMBER: US 60/290,029  
PRIOR FILING DATE: 2001-05-11  
PRIOR APPLICATION NUMBER: US 60/290,645  
PRIOR FILING DATE: 2001-05-15  
PRIOR APPLICATION NUMBER: US 60/292,336  
PRIOR FILING DATE: 2001-05-22  
PRIOR APPLICATION NUMBER: US 60/295,798  
PRIOR FILING DATE: 2001-06-06  
PRIOR APPLICATION NUMBER: US 60/297,457  
PRIOR FILING DATE: 2001-06-13  
PRIOR APPLICATION NUMBER: US 60/298,884  
PRIOR FILING DATE: 2001-06-19  
PRIOR APPLICATION NUMBER: US 60/303,459  
PRIOR FILING DATE: 2001-07-09  
NUMBER OF SEQ ID NOS: 1740  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1604  
LENGTH: 6822  
TYPE: DNA  
ORGANISM: Rattus norvegicus  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM\_013119  
US-09-917-800A-1604

Query Match 85.0%; Score 17; DB 10; Length 6822;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTG 17  
Db 2351 GCTGGAGGATTCCTCTG 2367

## RESULT 3

US-09-815-242-581/c  
Sequence 581, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 581  
LENGTH: 306  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-815-242-581

Query Match 79.0%; Score 15.8; DB 10; Length 306;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
Db 229 GCTGGAGGATTCATCTGTC 211

## RESULT 4

US-09-815-242-473/c  
Sequence 473, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/205,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 473  
LENGTH: 311  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-09-815-242-473

Query Match 79.0%; Score 15.8; DB 10; Length 311;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
Db 234 GCTGGAGGATTCATCTGTC 216

## RESULT 5

US-09-815-242-670/c  
Sequence 670, Application US/09815242

```
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 670
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-670

Query Match 79.0%; Score 15.8; DB 10; Length 311;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCTGGAGGATTCCTCTGTC 19
Db 234 GCTGGAGGATTCATCTGTC 216

RESULT 6
US-09-815-242-796/c
; Sequence 796, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 796
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-815-242-796

Query Match 79.0%; Score 15.8; DB 10; Length 311;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCTGGAGGATTCCTCTGTC 19
Db 234 GCTGGAGGATTCATCTGTC 216

RESULT 7
US-09-864-761-13156
; Sequence 13156, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 13156  
LENGTH: 533  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC021401.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.42  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45  
US-09-864-761-13156

Query Match 79.0%; Score 15.8; DB 10; Length 533;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||| ||||| |||||  
Db 278 GCTGGAGGATTCACAGTC 296

## RESULT 8

US-09-815-242-3834  
Sequence 3834, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3834  
LENGTH: 639

TYPE: DNA  
ORGANISM: Enterococcus faecalis

US-09-815-242-3834

Query Match 79.0%; Score 15.8; DB 10; Length 639;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||| ||||| |||||  
Db 373 GCTGGAGGATTCATCTGTC 391

## RESULT 9

US-09-815-242-6439  
Sequence 6439, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242  
PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 6439  
LENGTH: 696

TYPE: DNA  
ORGANISM: Enterococcus faecalis

NAME/KEY: CDS  
LOCATION: (1)...(696)

US-09-815-242-6439

Query Match 79.0%; Score 15.8; DB 10; Length 696;  
Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||| ||||| |||||  
Db 418 GCTGGAGGATTCATCTGTC 436

## RESULT 10

US-09-798-889-21  
Sequence 21, Application US/09798889  
Publication No. US20030004324A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: 31 Human secreted proteins

FILE REFERENCE: P2026P1  
CURRENT APPLICATION NUMBER: US/09/798,889

CURRENT FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022

PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-09  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714

PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686

PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,687

PRIOR FILING DATE: EARLIER FILING DATE: 1998-03-12  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,696

NUMBER OF SEQ ID NOS: 185  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 21  
LENGTH: 1727

TYPE: DNA

;; PRIOR FILING DATE: 2000-09-26

;	NUMBER OF SEQUENCES	CORRESPONDENCE
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

Steven Barash  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptic  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14

## ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

## INFORMATION FOR SEQ ID NO: 367:

SEQUENCE CHARACTERISTICS:  
LENGTH: 14557 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 367:  
US-09-070-927A-367

Query Match 79.0%; Score 15.8; DB 10; Length 14557;  
Best Local Similarity 89.5%; Pred. No. 1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||| ||||| |||||  
Db 3202 GCTGGAGGATTCATCTGTC 3184

## RESULT 15

US-09-782-378A-12  
; Sequence 12, Application US/09782378A  
; Patent No. US20020102731A1  
; GENERAL INFORMATION:  
; APPLICANT: Hearing, Patrick  
; APPLICANT: Bahou, Wadie  
; APPLICANT: Sandalou, Ziv  
; APPLICANT: Gnatenko, Dmitri  
; TITLE OF INVENTION: Adenoviral Vectors  
; FILE REFERENCE: STONYB-04970  
; CURRENT APPLICATION NUMBER: US/09/782,378A  
; PRIOR FILING DATE: 2001-02-12  
; PRIOR APPLICATION NUMBER: 60/237,747  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 36741  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-782-378A-12

Query Match 79.0%; Score 15.8; DB 10; Length 36741;  
Best Local Similarity 89.5%; Pred. No. 95;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19  
||||| ||||| |||||  
Db 10280 GCTGGAGGATTCCTCTGTC 10298

Search completed: July 9, 2003, 02:21:52  
Job time : 116.562 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1013.95 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-7  
Perfect score: 20  
Sequence: 1 gctggaggattctctgtcc 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estlin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_esti:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pin:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	92.0	273	14	BQ303884
2	18.4	92.0	793	17	BH486917
3	18	90.0	211	17	AZ379227
4	17.4	87.0	137	17	AZ567376
5	17.4	87.0	137	17	AZ567427
6	17.4	87.0	448	17	AQ783671

c	7	17.4	87.0	513	17	BH537780
	8	17.4	87.0	623	17	BH488309
	9	17.4	87.0	630	17	BH524153
c	10	17.4	87.0	760	17	BH672363
c	11	17.4	87.0	771	17	BH501851
c	12	17.4	87.0	804	17	BH716051
c	13	17.4	87.0	877	17	BH466014
c	14	16.8	84.0	315	12	BF840282
c	15	16.8	84.0	316	9	AV027372
c	16	16.8	84.0	322	12	BF840278
c	17	16.8	84.0	348	9	AI170426
c	18	16.8	84.0	351	12	BF355285
c	19	16.8	84.0	357	17	AZ499695
c	20	16.8	84.0	358	9	AA437457
c	21	16.8	84.0	373	10	AW238059
c	22	16.8	84.0	402	12	BF929991
c	23	16.8	84.0	437	17	B55831
c	24	16.8	84.0	449	9	AL630827
c	25	16.8	84.0	469	9	AI645117
c	26	16.8	84.0	472	17	AQ224410
c	27	16.8	84.0	502	9	AA798032
c	28	16.8	84.0	507	12	BF406708
c	29	16.8	84.0	510	17	AZ330954
c	30	16.8	84.0	526	9	AA174981
c	31	16.8	84.0	538	12	BF080527
c	32	16.8	84.0	541	17	B64751
c	33	16.8	84.0	549	10	BE445764
c	34	16.8	84.0	561	12	BE725072
c	35	16.8	84.0	579	9	AU045675
c	36	16.8	84.0	595	17	BH615514
c	37	16.8	84.0	598	10	AW211215
c	38	16.8	84.0	599	10	AW916963
c	39	16.8	84.0	609	17	BH764882
c	40	16.8	84.0	620	9	AV241470
c	41	16.8	84.0	622	17	AZ282935
c	42	16.8	84.0	641	13	BJ277239
c	43	16.8	84.0	650	10	BE427659
c	44	16.8	84.0	688	14	BQ240549
c	45	16.8	84.0	729	13	BI145864

#### ALIGNMENTS

RESULT 1  
BQ303884  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

BQ303884 273 bp mRNA linear EST 16-MAY-2002  
PM1-BT0759-200700-006-b07 BT0759 Homo sapiens cDNA, mRNA sequence.  
BQ303884  
BQ303884.1 GI:20821027  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 273)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldmab,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

```

Best Local Similarity 95.0%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCTCTGTCC 20
    |||||
Db 349 GCTGGAGGATTCTCTTTCC 368

```

RESULT 3  
AZ379227

AZ379227	211 bp	DNA	linear	GSS 02-OCT-2000
1M0134015F	10kb	plasmid	UUGC1M library	Mus musculus genomic
clone UUGC1M0134015 F				DNA sequence.
AZ379227				
AZ379227.1	GI:10492927			
GSS.				
house mouse.				
Mus musculus.				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
1 (bases 1 to 211)				

## REFERENCE AUTHORS

**TITLE**  
Mose, M., Rose, R., Stokes, R., Linney, A., von Niederhausern, A., and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL  
COMMENT

041127, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0134 row: 0 column: 15  
Seq primer: CGTTGTAACACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 211.  
Location/Qualifiers

```

i. .211
source
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="TUGC1M0134015"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to

```

electroporation. Vector DNA was prepared from a derivative of pWD42 (g114732114[gb]AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

### Query Match

90.0%; Score 18; DB 17; Length 211;

Best Local Similarity 100.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGT 18  
 Db 170 GCTGGAGGATTCCTCTGT 187

## RESULT 4

AZ567376 137 bp DNA linear GSS 07-MAY-2001  
 LOCUS 234PVA05 PV MBN #26/#27 (amplified once) Plasmodium vivax genomic  
 DEFINITION 3', DNA sequence.

ACCESSION AZ567376 GI:13975412

VERSION AZ567376

KEYWORDS GSS.

SOURCE malaria parasite P. vivax.

ORGANISM Plasmodium vivax

REFERENCE 1 (bases 1 to 137)

AUTHORS Carlton,J.M.-R. and Dame,J.B.

TITLE The Plasmodium vivax and P. Berghel gene sequence tag projects

JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)

COMMENT Contact: Dame JB

Dept. of Pathobiology, College of Veterinary Medicine

University of Florida

2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA

Tel: 352 392 4700

Fax: 352 392 9704

Email: damej@mail.vetmed.ufl.edu

Seq primer: M13(-20) forward

Class: shotgun.

Location/Qualifiers

source

1. 137

/organism="Plasmodium vivax"

/strain="Belem"

/db\_xref="taxon:5855"

/clone\_lib="PV MBN #26/#27 (amplified once)"

/dev\_stage="asexual blood forms"

/lab\_host="Saimiri boliviensis"

/note="Vector: Lambda ZAP II (Stratagene); individual

clones excised into phagemid pBluescript; Site 1: EcoR I;

Site 2: EcoR I; Genomic DNA was prepared from asynchronous

blood stage forms of the Belem line of P. vivax grown in

squirrel monkeys. Parasitized erythrocytes were purified

from contaminating host leukocytes by filtration of ADP

activated blood through acid-washed glass beads and

Whatman Cfl1 cellulose columns by gravity filtration.

Purified DNA was digested with mung bean nuclease in the

presence of 42.5% formamide at 500C as described (Galinski

, M. et al. 1992. Cell 69,1213-1226; Vernick, K.D. et

al.1988. N.A.R. 16, 6883-6896). Eco RI linkers were added

and the constructs ligated into Lambda ZAP II. P. vivax

Belem was originally isolated from a patient in Belem,

Brazil 1980 by Mercia de Arruda, adapted to Saimiri

monkeys by Jurg Gysin, and maintained since 1983 in

squirrel monkeys.

BASE COUNT 46 a 26 c 23 g 42 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 137;

Best Local Similarity 94.7%; Pred. No. 4.4e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CTGGAGGATTCCTCTGTCC 20

Db 111 CTGGAGGATTCCTCTGTCC 129

RESULT 5

AZ567427 137 bp DNA linear GSS 07-MAY-2001

LOCUS 234PVF01 PV MBN #26/#27 (amplified once) Plasmodium vivax genomic

DEFINITION 3', DNA sequence.

ACCESSION AZ567427 GI:13975516

VERSION AZ567427

KEYWORDS GSS.

SOURCE malaria parasite P. vivax.

ORGANISM Plasmodium vivax

REFERENCE 1 (bases 1 to 137)

AUTHORS Carlton,J.M.-R. and Dame,J.B.

TITLE The Plasmodium vivax and P. Berghel gene sequence tag projects

JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)

COMMENT Contact: Dame JB

Dept. of Pathobiology, College of Veterinary Medicine

University of Florida

2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA

Tel: 352 392 4700

Fax: 352 392 9704

Email: damej@mail.vetmed.ufl.edu

Seq primer: M13(-20) forward

Class: shotgun.

Location/Qualifiers

source

1. 137

/organism="Plasmodium vivax"

/strain="Belem"

/db\_xref="taxon:5855"

/clone\_lib="PV MBN #26/#27 (amplified once)"

/dev\_stage="asexual blood forms"

/lab\_host="Saimiri boliviensis"

/note="Vector: Lambda ZAP II (Stratagene); individual

clones excised into phagemid pBluescript; Site 1: EcoR I;

Site 2: EcoR I; Genomic DNA was prepared from asynchronous

blood stage forms of the Belem line of P. vivax grown in

squirrel monkeys. Parasitized erythrocytes were purified

from contaminating host leukocytes by filtration of ADP

activated blood through acid-washed glass beads and

Whatman Cfl1 cellulose columns by gravity filtration.

Purified DNA was digested with mung bean nuclease in the

presence of 42.5% formamide at 500C as described (Galinski

, M. et al. 1992. Cell 69,1213-1226; Vernick, K.D. et

al.1988. N.A.R. 16, 6883-6896). Eco RI linkers were added

and the constructs ligated into Lambda ZAP II. P. vivax

Belem was originally isolated from a patient in Belem,

Brazil 1980 by Mercia de Arruda, adapted to Saimiri

monkeys by Jurg Gysin, and maintained since 1983 in

squirrel monkeys.

BASE COUNT 46 a 26 c 23 g 42 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 137;

Best Local Similarity 94.7%; Pred. No. 4.4e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CTGGAGGATTCCTCTGTCC 20

Db 111 CTGGAGGATTCCTCTGTCC 129

RESULT 6

AZ567427 137 bp DNA linear GSS 03-AUG-1999

LOCUS HS\_3098\_B1\_D05\_MR CIT Approved Human Genomic Sperm Library D Homo

DEFINITION sapiens genomic clone.Plate-3098 Col-9 Row-R, Homo

ACCESSION AO783671

VERSION AO783671

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 448)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 448)

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and

Hood,L.

3', DNA sequence.  
 ACCESSION AZ567427 GI:13975516  
 VERSION AZ567427.1  
 KEYWORDS GSS.  
 SOURCE malaria parasite P. vivax.  
 ORGANISM Plasmodium vivax  
 REFERENCE 1 (bases 1 to 137)  
 AUTHORS Carlton,J.M.-R. and Dame,J.B.  
 TITLE The Plasmodium vivax and P. Berghel gene sequence tag projects  
 JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)  
 COMMENT Contact: Dame JB  
 Dept. of Pathobiology, College of Veterinary Medicine  
 University of Florida  
 2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA  
 Tel: 352 392 4700  
 Fax: 352 392 9704  
 Email: damej@mail.vetmed.ufl.edu  
 Seq primer: M13(-20) forward  
 Class: shotgun.  
 Location/Qualifiers  
 1. 137  
 /organism="Plasmodium vivax"  
 /strain="Belem"  
 /db\_xref="taxon:5855"  
 /clone\_lib="PV MBN #26/#27 (amplified once)"  
 /dev\_stage="asexual blood forms"  
 /lab\_host="Saimiri boliviensis"  
 /note="Vector: Lambda ZAP II (Stratagene); individual  
 clones excised into phagemid pBluescript; Site 1: EcoR I;  
 Site 2: EcoR I; Genomic DNA was prepared from asynchronous  
 blood stage forms of the Belem line of P. vivax grown in  
 squirrel monkeys. Parasitized erythrocytes were purified  
 from contaminating host leukocytes by filtration of ADP  
 activated blood through acid-washed glass beads and  
 Whatman Cfl1 cellulose columns by gravity filtration.  
 Purified DNA was digested with mung bean nuclease in the  
 presence of 42.5% formamide at 500C as described (Galinski  
 , M. et al. 1992. Cell 69,1213-1226; Vernick, K.D. et  
 al.1988. N.A.R. 16, 6883-6896). Eco RI linkers were added  
 and the constructs ligated into Lambda ZAP II. P. vivax  
 Belem was originally isolated from a patient in Belem,  
 Brazil 1980 by Mercia de Arruda, adapted to Saimiri  
 monkeys by Jurg Gysin, and maintained since 1983 in  
 squirrel monkeys.

BASE COUNT 46 a 26 c 23 g 42 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 137;  
 Best Local Similarity 94.7%; Pred. No. 4.4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CTGGAGGATTCCTCTGTCC 20  
 Db 111 CTGGAGGATTCCTCTGTCC 129

RESULT 6  
 AO783671/c 448 bp DNA linear GSS 03-AUG-1999  
 LOCUS HS\_3098\_B1\_D05\_MR CIT Approved Human Genomic Sperm Library D Homo  
 DEFINITION sapiens genomic clone.Plate-3098 Col-9 Row-R, Homo  
 ACCESSION AO783671  
 VERSION AO783671  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 448)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 448)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC end web server: http://www.htsc.washington.edu  
 Plate: 3098 row: H column: 9  
 Seq primer: M13 Reverse  
 Class: BAC ends  
 High quality sequence stop: 448.

## FEATURES

source  
 1. .448  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="plate-3098 Col-9 Row-H"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"  
 BASE COUNT 117 a 86 c 99 g 146 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 448;  
 Best Local Similarity 94.7%; Pred. No. 6.6e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 CTGGAGGATTCCTCTGCC 20  
 Db 344 CTGGAGGATTCCTCTGCC 326

RESULT 7  
 BH537780/c  
 LOCUS  
 DEFINITION BOGNB38TR BOGN Brassica oleracea genomic clone BOGNB38, DNA  
 ACCESSION BH537780  
 VERSION BH537780.1 GI:17775365  
 KEYWORDS  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 513)  
 Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other GSSs: BOGNB38TF  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.  
 Location/Qualifiers  
 1. .513  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGNB38"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 122 a 135 c 147 g 219 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 623;  
 Best Local Similarity 94.7%; Pred. No. 7.4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 19  
 Db 493 GCTGGAGGATTCCTCTTTC 511

## FEATURES

source  
 1. .513  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGNB38"  
 /clone\_lib="BOGN"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 162 a 144 c 137 g 70 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 513;  
 Best Local Similarity 94.7%; Pred. No. 7e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 19  
 Db 480 GCTGGAGGATTCCTCTTTC 462

RESULT 8  
 BH488309  
 LOCUS  
 DEFINITION BOGTG10TF BOGT Brassica oleracea genomic clone BOGTG10, DNA  
 ACCESSION BH488309  
 VERSION BH488309.1 GI:17696413  
 KEYWORDS  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 623)  
 Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other GSSs: BOGTG10TR  
 Contact: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TF  
 Class: sheared ends.  
 Location/Qualifiers  
 1. .623  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGTG10"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 122 a 135 c 147 g 219 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 623;  
 Best Local Similarity 94.7%; Pred. No. 7.4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 19  
 Db 493 GCTGGAGGATTCCTCTTTC 511

## FEATURES

source  
 1. .623  
 /organism="Brassica oleracea"  
 /strain="TO1000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOGTG10"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 122 a 135 c 147 g 219 t  
 ORIGIN  
 Query Match 87.0%; Score 17.4; DB 17; Length 623;  
 Best Local Similarity 94.7%; Pred. No. 7.4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GCTGGAGGATTCCTCTGTC 19  
 Db 493 GCTGGAGGATTCCTCTTTC 511

RESULT 9  
 BH524153  
 LOCUS  
 DEFINITION BOHBY07TR BOHB Brassica oleracea genomic clone BOHBY07, DNA  
 ACCESSION BH524153  
 VERSION BH524153.1 GI:17732238  
 KEYWORDS  
 SOURCE Brassica oleracea.  
 ORGANISM Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae: eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 630)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOHBY07F  
Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. 630  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOHBY07"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

127 a 128 c 156 g 219 t

BASE COUNT  
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 630;  
Best Local Similarity 94.7%; Pred. No. 7.5e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19

Db 489 GCTGGAGGATTCCTCTTC 507

RESULT 10  
BH672363/c

LOCUS  
DEFINITION  
BOMMT65TF BO\_2\_3\_KB Brassica oleracea genomic clone BOMMT65, DNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

Brassica oleracea.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 760)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. 760  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOMMT65"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

247 a 198 c 173 g 142 t

BASE COUNT  
ORIGIN

Query Match

87.0%; Score 17.4; DB 17; Length 760;  
Best Local Similarity 94.7%; Pred. No. 8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19

Db 431 GCTGGAGGATTCCTCTTC 413

RESULT 11  
BH501851/c

LOCUS  
DEFINITION  
BOMKW15TR BOGK Brassica oleracea genomic clone BOGKW15, DNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 771)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
Whole genome shotgun sequencing of Brassica oleracea  
Unpublished (2001)  
Other GSSs: BOGKW15TF  
Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. 771  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOGKW15"  
/clone\_lib="BOGK"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

240 a 198 c 192 g 141 t

BASE COUNT

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 771;  
Best Local Similarity 94.7%; Pred. No. 8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19

Db 458 GCTGGAGGATTCCTCTTC 440

RESULT 12  
BH716051/c

LOCUS  
DEFINITION  
BOMEW59TR BO\_2\_3\_KB Brassica oleracea genomic clone BOMEW59, DNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 804)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. 804  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOMEW59"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

804 bp DNA linear GSS 20-FEB-2002

BASE COUNT

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 804;  
Best Local Similarity 94.7%; Pred. No. 8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTGTC 19

Db 458 GCTGGAGGATTCCTCTTC 440

RESULT 12  
BH716051/c

LOCUS  
DEFINITION  
BOMEW59TR BO\_2\_3\_KB Brassica oleracea genomic clone BOMEW59, DNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM

Brassica oleracea.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 804)  
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TF

Class: sheared ends.

Location/Qualifiers

1. 804  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOMEW59"  
/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared  
genomic DNA inserted into PHOS1 using BstXI linkers"

804 bp DNA linear GSS 20-FEB-2002

BASE COUNT

ORIGIN

TITLE Whole genome shotgun sequencing of Brassica oleracea  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: BOMEW59TF  
Contact: Chris Town

TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208

Email: cdtown@tigr.org  
DNA is from a doubled haploid provided by Tom Osborn.  
Seq primer: TR  
Class: sheared ends.

Location/Qualifiers

1. 804  
/organism="Brassica oleracea"  
/strain="TO1000DH3"  
/db\_xref="taxon:3712"  
/clone="BOMEW59"  
/note="Vector: phos1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into phos1 using BstXI linkers"

278 a 194 c 179 g 153 t

Query Match 87.0%; Score 17.4; DB 17; Length 804;  
Best Local Similarity 94.7%; Pred. No. 8.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTCTC 19  
|||||  
Db 272 GCTGGAGGATTCCTCTTTC 254.

RESULT 13  
BH466014/c  
LOCUS BOHRJ30TR BOHR Brassica oleracea genomic clone BOHRJ30, DNA  
DEFINITION sequence.  
ACCESSION BH466014  
VERSION BH466014.1 GI:17665580  
KEYWORDS GSS.  
SOURCE Brassica oleracea.  
ORGANISM Brassica oleracea

Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 877)

Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other\_GSSs: BOHRJ30TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. 877

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHRJ30"

/note="Vector: phos1; Site\_1: BstXI; 2-3 kb sheared genomic DNA inserted into phos1 using BstXI linkers"

275 a 234 c 202 g 166 t

Query Match 87.0%; Score 17.4; DB 17; Length 877;

AV027372/c

LOCUS

AV027372

316 bp mRNA linear EST 31-AUG-1999

Best Local Similarity 94.7%; Pred. No. 8.4e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTCTC 19  
|||||  
Db 443 GCTGGAGGATTCCTCTTTC 425

RESULT 14  
BF840282  
LOCUS RC4-HT0986-291100-011-h02 HT0986 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF840282  
ACCESSION BF840282  
VERSION BF840282.1 GI:12192745  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 315)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC4&t2=RC4-HT0986-  
291100-011-h02&t3=2000-11-29&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 51

High quality sequence stop: 315.

Location/Qualifiers

1. 315

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0986"

/dev\_stage="Adult"

/note="Organ: head; neck; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

46 a 76 c 108 g 85 t

Query Match 84.0%; Score 16.8; DB 12; Length 315;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCTGGAGGATTCCTCTCTC 20  
|||||

Db 69 GCTGGAGGATTCCTCTCTC 88

RESULT 15  
AV027372/c

LOCUS

AV027372

316 bp mRNA linear EST 31-AUG-1999

Search completed: July 8, 2003, 09:21:21  
Job time : 1019.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53; Search time 209,093 seconds  
(without alignments)  
2644,537 Million cell updates/sec

Title: US-09-647-780A-8  
Perfect score: 19  
Sequence: 1 cggggatcacatggtgcg 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vi: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_ov: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vi: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rod: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrt: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	19	100.0	19	6	AX014708 Sequence
2	19	100.0	21	6	AX014710 Sequence
3	19	100.0	2765	6	AX014701 Sequence
4	19	100.0	174953	2	AC094732 Rattus no
5	16.4	86.3	37476	9	AL0316612 Human DNA
6	16.4	86.3	173200	2	AC121922 Mus muscu
7	16.4	86.3	214042	2	AC093028 Homo sapi
8	16.4	86.3	279011	9	AE006467 Homo sapi
9	16.4	86.3	294250	1	AP001517 Bacillus
10	16	84.2	119790	9	HSJ989D7 Human DNA
11	16	84.2	160819	2	AC026171 Homo sapi
12	15.8	83.2	504	6	AR206865 Sequence
13	15.8	83.2	504	6	AX136894 Sequence
14	15.8	83.2	504	6	BD006820 N-Acetyl
15	15.8	83.2	1107	6	AR206864 Sequence
16	15.8	83.2	1107	6	AX136892 Sequence
17	15.8	83.2	1107	6	AX417395 Sequence
18	15.8	83.2	1107	6	AX427614 Sequence
19	15.8	83.2	1107	6	BD006819 N-Acetyl
20	15.8	83.2	1311	1	AOR292519 Anycolato
21	15.8	83.2	5500	1	AF153086 Burkholde
22	15.8	83.2	11036	1	AE012477 Xanthomon
23	15.8	83.2	12295	1	AE011745 Xanthomon
24	15.8	83.2	44381	2	AC097605 Rattus no
25	15.8	83.2	97348	2	AC102958 Rattus no
26	15.8	83.2	99304	9	AC106825 Homo sapi
27	15.8	83.2	126582	9	AC1010241 Homo sapi
28	15.8	83.2	142970	2	AL355811 Homo sapi
29	15.8	83.2	150891	2	AC069098 Homo sapi
30	15.8	83.2	155585	9	AL353671 Human DNA
31	15.8	83.2	160093	2	AC109770 Rattus no
32	15.8	83.2	178348	9	AP000501 Homo sapi
33	15.8	83.2	180722	2	AC105393 Homo sapi
34	15.8	83.2	185069	2	AC084692 Homo sapi
35	15.8	83.2	188711	9	AC093326 Homo sapi
36	15.8	83.2	198084	2	AC019339 Homo sapi
37	15.8	83.2	200242	2	AC115037 Mus muscu
38	15.8	83.2	349498	1	AP003002 Mesorhizo
39	15.4	81.1	633	3	AF139520 Caenorhab
40	15.4	81.1	2494	10	BC024640 Mus muscu
41	15.4	81.1	4063	10	GPINCP1B2 Guinea pig
42	15.4	81.1	4082	10	GPINCP1A1 Guinea pig
43	15.4	81.1	5162	3	AY069786 Drosophil
44	15.4	81.1	29861	3	AF000298 Caenorhab
45	15.4	81.1	43632	9	AC108358 Homo sapi

ALIGNMENTS

RESULT 1  
AX014708  
LOCUS AX014708 19 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 8 from Patent WO9553077.  
ACCESSION AX014708  
VERSION AX014708.1 GI:10040981  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial construct  
REFERENCE 1 (bases 1 to 19)  
AUTHORS Schwartz J.C., Gros C., Ouimet T., Rose C., Bonhomme M.C. and Facchinetti P.  
TITLE Novel nep II membrane metalloprotease and its use for screening inhibitors useful in therapy

inhibitors useful in therapy  
Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANIT RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUIMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES  
source Location/Qualifiers  
1. .2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107 2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSSSSVGMERADNCGRRRLRGFVECGLLVLTLLLMCAIVTL  
GVFYSIGKOLPILNLSLLVSHRRTVVKVLKSDSKSDICTPSCVIAARATLQNMV  
QSKKPCDNFYQACGSLWLRHHVYPETNSYSVPFDILRDEVLILKGLVLESDSSVQHRPA  
VERAKNTLYRSCNMGVSRKEDRSEPLNLNDMIGMVPKADKNWNETMGPKEWELERQLRA  
LNSQFNRRVLLDLFWNDONSRRHYIIDQTLGMPSEYFYKEDSHRVREAYLQFM  
TSVATMLRDLLNPGETDLVQEMAOVLHLETHLANATVPQKRHDVTALYHRMGLEE  
LORFGLKGFNFTLFTONVLSVQVLLNEEVVYGIPLNLEIETIDVFPATQNLN  
YLWRLVLDRIQSLSORFEARVDYRKALYGTMEVRWRKCVSYVNSNSAVGSLY  
IKRASKDSKSIYSELLEKIRSVFVDNLDELNMDESKKKAQEKALNIREQIGPYD  
ILEDNNRHDEYSLSFTFSEFNGLNKNNAORSUKLRKRVQDQNLWIGAAV  
NIFYSPNNRLVFPAGILPPFPFSDQPOLNFGGIGMWIGHEITGFPDNGRNGVQA  
GNMLWDSNFSARHFRQSQCMIIQYSNFSELAQNVNGFSTLGENIADNGGVQA  
YKAYLQWLAEAGRDRLQNLNTYAQLFTINVAQVWCGSYRPEFAIQSTKTDVHSLPK  
YRVLSGLNLPGFSEAFHCPGSRPHMPCRIV"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 19; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGGGGATCACATGGTGCCG 19  
|||||  
Db 485 CGGGGATCACATGGTGCCG 467  
RESULT 4  
AC094732  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-516, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
56 unordered pieces.  
AC094732  
AC094732.2 GI:17941511  
HTG: HTGS\_PHASE1.  
Rattus norvegicus.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 174953)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amarating,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Grantz,P., Gabis,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Haves,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,H., King,I., Korvah,I.,

Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
 Louseged, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
 Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,  
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M. P., Meador, M.,  
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,  
 Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
 Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,  
 Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R.,  
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
 Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,  
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Unpublished  
 2 (bases 1 to 174953)

Worley, K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:15624568.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: BGFG

Center clone name: CH230-516

----- Summary Statistics

Assembly program: Phrap; version 0.9903295first call to

findPhrapList

Consensus quality: 152255 bases at least Q30

Consensus quality: 158448 bases at least Q40

Consensus quality: 164461 bases at least Q20

Estimated insert size: 155965; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 56 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 13782: contig of 13782 bp in length

\* 13783 13882: gap of unknown length

\* 13883 23287: contig of 9405 bp in length

\* 23288 23387: gap of unknown length

\* 23388 28081: contig of 4694 bp in length

\* 28082 28181: gap of unknown length

\* 28182 33807: contig of 5626 bp in length

\* 33808 33907: gap of unknown length

\* 33908 39271: contig of 5364 bp in length

\* 39272 39371: gap of unknown length

\* 39372 44270: contig of 4899 bp in length

\* 44271 44370: gap of unknown length

\* 44371 47723: contig of 3353 bp in length

\* 47724 47823: gap of unknown length

\* 47824 53427: contig of 5604 bp in length

\* 53428 53527: gap of unknown length

\* 53528 57303: contig of 3776 bp in length

\* 57304 57403: gap of unknown length

\* 57404 62018: contig of 4615 bp in length

\* 62019 62118: gap of unknown length

\* 62119 66676: contig of 4358 bp in length

\* 66677 66776: gap of unknown length

\* 66777 70201: contig of 3425 bp in length

\* 70202 70301: gap of unknown length

\* 70302 74082: contig of 3781 bp in length

\* 74083 74182: gap of unknown length

\* 74183 78306: contig of 4124 bp in length

\* 78307 78406: gap of unknown length

\* 78407 81222: contig of 2816 bp in length

\* 81223 81322: gap of unknown length

\* 81323 84552: contig of 3230 bp in length

\* 84553 84652: gap of unknown length

\* 84653 88839: contig of 4187 bp in length

\* 88840 88939: gap of unknown length

\* 88940 92256: contig of 3317 bp in length

\* 92257 92356: gap of unknown length

\* 92357 94995: contig of 2639 bp in length

\* 94996 95095: gap of unknown length

\* 95096 98465: contig of 3370 bp in length

\* 98466 98565: gap of unknown length

\* 98566 101632: contig of 3067 bp in length

\* 101633 101732: gap of unknown length

\* 101733 104939: contig of 3207 bp in length

\* 104940 105039: gap of unknown length

\* 105040 109248: contig of 4209 bp in length

\* 109249 109348: gap of unknown length

\* 109349 112435: contig of 3087 bp in length

\* 112436 112536: gap of unknown length

\* 112537 114676: contig of 2141 bp in length

\* 114677 114776: gap of unknown length

\* 114777 117567: contig of 2791 bp in length

\* 117568 117667: gap of unknown length

\* 117668 119209: contig of 1542 bp in length

\* 119210 121797: contig of 2488 bp in length

\* 121798 121897: gap of unknown length

\* 121898 124293: contig of 2396 bp in length

\* 124294 124393: gap of unknown length

\* 124394 126428: contig of 2035 bp in length

\* 126429 128528: gap of unknown length

\* 128529 129525: contig of 2997 bp in length

\* 129526 129626: gap of unknown length

\* 129627 132334: contig of 2709 bp in length

\* 132335 132434: gap of unknown length

\* 132435 135274: contig of 2840 bp in length

\* 135275 135374: gap of unknown length

\* 135375 138074: contig of 2700 bp in length

\* 138075 138174: gap of unknown length

\* 138175 139985: contig of 1811 bp in length

\* 139986 140085: gap of unknown length

\* 140086 142273: contig of 2188 bp in length

\* 142274 143598: contig of 1225 bp in length

\* 143599 143698: gap of unknown length

\* 143699 145434: contig of 1736 bp in length

\* 145435 145535: gap of unknown length

\* 145536 146985: contig of 1451 bp in length

\* 146986 147085: gap of unknown length

\* 147086 148099: contig of 1014 bp in length

\* 148100 148199: gap of unknown length

\* 148200 150915: contig of 2716 bp in length

\* 150916 151015: gap of unknown length

\* 151016 152501: contig of 1486 bp in length

\* 152502 152601: gap of unknown length

\* 152602 154010: contig of 1409 bp in length

\* 154011 154110: gap of unknown length

\* 154111 155758: contig of 1648 bp in length

\* 155759 155858: gap of unknown length

\* 155859 157622: contig of 1764 bp in length

```

* 157623 157722: gap of unknown length
* 157723 159428: contig of 1706 bp in length
* 159429 159528: gap of unknown length
* 159529 161209: contig of 1681 bp in length
* 161210 161309: gap of unknown length
* 161310 163413: contig of 2104 bp in length
* 163414 163513: gap of unknown length
* 163514 164702: contig of 1189 bp in length
* 164703 164802: gap of unknown length
* 164803 165998: contig of 1196 bp in length
* 165999 166098: gap of unknown length
* 166099 167412: contig of 1314 bp in length
* 167413 167512: gap of unknown length
* 167513 169231: contig of 1719 bp in length
* 169232 169331: gap of unknown length
* 169332 170534: contig of 1203 bp in length
* 170535 170634: gap of unknown length
* 170635 172047: contig of 1413 bp in length
* 172048 172147: gap of unknown length
* 172148 173509: contig of 1362 bp in length
* 173510 173610: gap of unknown length
* 173610 174953: contig of 1344 bp in length.

```

## FEATURES

Location/Qualifiers

```

Query Match      100.0%; Score 19; DB 2; Length 174953;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 CGGGATCACATGGTGGCG 19

```

|||||
Db 40463 CGGGATCACATGGTGGCG 40481

```

## RESULT 5

## HS316G12

## LOCUS

## DEFINITION

Human DNA sequence from clone LA16-316G12 on chromosome 16. Contains the gene for C2 domain protein KIAA0734, the gene for a novel protein similar to predicted yeast, worm and archaea-bacterial proteins, a novel gene and the 3' part of the gene for a novel protein similar to API1 and API2 (apoptosis inhibitor 1 and 2 (MIHB, MIHC, IAP1, IAP2)). Contains ESTs, a GSS and six putative CpG islands, complete sequence.

## ACCESSION

AL031709.12

GI:13937339

HTG; API1; API2; CpG island; IAP1; IAP2; KIAA0734; MIHB; MIHC.

human.

## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (Bases 1 to 37476)

Frankland.J.

Direct Submission

Submitted (19-Apr-2001) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequests@sanger.ac.uk

On May 3, 2001 this sequence version replaced gi:13277223.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations.

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality &gt;

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em; EMBL; Sw;

SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP

database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) LA16-316G12 is  
 part of a clone contig from the tip of the short arm of chromosome  
 16 spanning 2Mb of pl3.3 (Higgs D.R., Flint J., Daniels R., MRC  
 Molecular Haematology Unit, Institute of Molecular Medicine, Oxford  
 (unpublished)), and is from the Los Alamos, flow sorted human  
 Chromosome 16 libraries constructed by Norman Doggett  
 (unpublished). VECTOR: sCos-1  
 IMPORTANT: This sequence is not the entire insert of clone  
 LA16-316G12 it may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone LA16-316G12 is at 1 in this sequence.  
 The true left end of clone LA16-396E4 is at 37377 in this sequence.  
 The true right end of clone LA16-358B7 is at 21289 in this  
 sequence.

## FEATURES

source

1. 37476

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="16"

/clone="LA16-316G12"

/clone\_lib="LA16"

987. 2805

/note="CpG island"

/evidence=not\_experimental

join(1727. 1820,6611. 6751,6973. 7060,7235. 7315,

7567. 7674,8993. 9052,9200. 9266,9365. 9560,9893. 9937,

10057. 10134,10252. 10354,10598. 10705,10790. 10885,

10984. 11087,11464. 11584,12093. 12217,12388. 12388,

12540. 12596,12672. 12769,12869. 12920,13040. 13170,

13320. 13443,13818. 13961,14057. 14149,14241. 14432,

14513. 14581,14676. 14755,14918. 15026,15369. 15440,

15616. 15696,15772. 15892,15965. 16088,16167. 16372,

16450. 17510)

/gene="c316G12.1"

join(1727. 1820,6611. 6751,6973. 7060,7235. 7315,

7567. 7674,8993. 9052,9200. 9266,9365. 9560,9893. 9937,

10057. 10134,10252. 10354,10598. 10705,10790. 10885,

10984. 11087,11464. 11584,12093. 12217,12388. 12388,

12540. 12596,12672. 12769,12869. 12920,13040. 13170,

13320. 13443,13818. 13961,14057. 14149,14241. 14432,

14513. 14581,14676. 14755,14918. 15026,15369. 15440,

15616. 15696,15772. 15892,15965. 16088,16167. 16372,

16450. 17510)

/gene="c316G12.1"

/product="c316G12.1 (KIAA0734 (C2 domain protein))"

/note="match: CDNAS: Em:AB018277 Em:AF038191

match: ESTs: Em:T77480 Em:H18551 Em:F13300 Em:R61025

Em:R24790 Em:R85179 Em:A1217078 Em:AA078749 Em:A1696810

Em:A1086907 Em:H15802 Em:AA293357 Em:H46334 Em:D81844

Em:A1301780 Em:R45522 Em:AA782001 Em:R60906 Em:AA716721

Em:A1613004 Em:AA716524 Em:C00805 Em:A111375 Em:AA481839

Em:W01878 Em:A1146703 Em:AA627545 Em:AA079823 Em:A1240952

Em:AA258690 Em:AA482458 Em:AA682952 Em:A1638343

Em:A1693236 Em:A1537389 Em:H18444 Em:A158964 Em:A1564977

Em:AA302073"

/evidence=not\_experimental

join(16621. 6751,6973. 7060,7235. 7315,7567. 7674,

8993. 9052,9200. 9266,9365. 9560,9893. 9937,10057. 10134,

10252. 10354,10598. 10705,10790. 10885,10984. 11087,

11464. 11584,12093. 12217,12388. 12388,12540. 12596,

12672. 12769,12869. 12920,13040. 13170,13320. 13443,

13818. 13961,14057. 14149,14241. 14432,14513. 14581,

14676. 14755,14918. 15026,15369. 15440,15616. 15696,

15772. 15892,15965. 16088,16167. 16372,16450. 16553)

/gene="c316G12.1"

/codon\_start=1

/evidence=not\_experimental

/product="c316G12.1 (KIAA0734 (C2 domain protein))"

/protein\_id="CAB56182.1"

/db\_xref="GI:5912546"

/translation="MSTLLDIKSSVLKQVQVCFRRRTEDQDPCGSADPOEPATGAW

RKDGVEFFAHRLMLKKGEGKGLPCLEVLKSGSPAPPEPVPDSLGLRALAPEVE



\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 654: contig of 654 bp in length  
655 754: gap of unknown length  
755 1280: contig of 1226 bp in length  
2081 2080: gap of unknown length  
3271 3271: contig of 1191 bp in length  
3371 3371: gap of unknown length  
3372 7512: contig of 4141 bp in length  
7513 7612: gap of unknown length  
7613 11643: contig of 4031 bp in length  
11644 11743: gap of unknown length  
11744 25006: contig of 13263 bp in length  
25007 25106: gap of unknown length  
25107 45336: contig of 20430 bp in length  
45337 45636: gap of unknown length  
45637 69889: contig of 24253 bp in length  
69890 69889: gap of unknown length  
69890 100569: contig of 30580 bp in length  
100570 100669: gap of unknown length  
100670 147725: contig of 47056 bp in length  
147726 147825: gap of unknown length  
147826 173200: contig of 25375 bp in length.

FEATURES

source  
1...173200  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="UNK"  
/clone="RP24-198H19"  
1..654  
/note="assembly\_name:Contig12"  
755..1980  
/note="assembly\_name:Contig14"  
2081..3271  
/note="assembly\_name:Contig15"  
3372..7512  
/note="assembly\_name:Contig16"  
7613..11643  
/note="assembly\_name:Contig17"  
11744..25006  
/note="assembly\_name:Contig18"  
25107..45336  
/note="assembly\_name:Contig19"  
45637..69889  
/note="assembly\_name:Contig20"  
69990..100569  
/note="assembly\_name:Contig21"  
100670..147725  
/note="assembly\_name:Contig22"  
147826..173200  
/note="assembly\_name:Contig23"  
BASE COUNT 43220 a 44056 c 43959 g 40962 t 1003 others  
ORIGIN

Query Match 86.3%; Score 16.4; DB 2; Length 173200;  
Best Local Similarity 94.4%; Pred. NO. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGTCGCC 18  
Db 15610 CAGGGATCACATGTCGCC 15627

RESULT 7

AC093028 AC093028 214042 bp DNA linear HTG 31-JUL-2002  
DEFINITION Homo sapiens chromosome 12 clone RP11-604N13, WORKING DRAFT  
SEQUENCE, 3 unordered pieces.  
AC093028  
VERSION AC093028.9 GI:21954889

KEYWORDS  
SOURCE  
ORGANISM

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 214042)

REFERENCE  
AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsnbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
Bouch,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisege,H.,  
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,N.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,  
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Umsani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

TITLE  
JOURNAL

2 (bases 1 to 214042)  
Unpublished  
Worley,K.C.

REFERENCE  
AUTHORS

Direct Submission  
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 214042)  
Worley,K.C.

REFERENCE  
AUTHORS

Direct Submission  
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 25, 2002 this sequence version replaced gi:21553148.

COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Drafting Center Code: BCM  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: HDJK  
Center clone name: RP11-604N13  
Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads

```

Assembly program: Phrap; version 0.990329
Consensus quality: 213137 bases at least Q40
Consensus quality: 213777 bases at least Q30
Consensus quality: 214432 bases at least Q20
Estimated insert size: 211386; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 3 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
*
1 2660: contig of 2660 bp in length
*
2661 2760: gap of unknown length
*
2761 65406: contig of 62646 bp in length
*
65407 65506: gap of unknown length
*
65507 214042: contig of 148536 bp in length.
*
Location/Qualifiers
1. 214042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-604N13"
BASE COUNT 53357 a 50252 c 51683 g 57749 t 1001 others
ORIGIN
Query Match 86.3%; Score 16.4; DB 2; Length 214042;
Best Local Similarity 94.4%; Pred. No. 2.9e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGGATCACATGTCGCG 19
Db 89130 GGTGATCACATGTCGCG 89147

RESULT 8
AE006467 279011 bp DNA linear PRI 15-AUG-2002
LOCUS Homo sapiens 16p13.3 sequence section 6 of 8.
DEFINITION AE006467 AE005175
ACCESSION AE006467.1 GI:143336745
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 279011)
AUTHORS Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,
Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and
Higgs,D.R.
Sequence, structure and pathology of the fully annotated terminal 2
Mb of the short arm of human chromosome 16
Hum. Mol. Genet. 10 (4), 339-352 (2001)
JOURNAL 21096910
MEDLINE
PUBMED 11157797
REFERENCE 2 (bases 1 to 279011)
AUTHORS Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,
Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and
Higgs,D.R.
Direct Submission
TITLE Submitted (07-DEC-2000) MRC Molecular Haematology Unit, Weatherall
JOURNAL Institute of Molecular Medicine, John Radcliffe Hospital, Oxford,
Oxon OX3 9DS, UK
FEATURES
Location/Qualifiers
1. 279011
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"

```

```

/map="16p13.3"
1. 14033
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 358b7"
1833. 46356
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 316g12"
39209..78044
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 399e4"
75764. 110196
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 312e8"
95097. 126950
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 438f12"
110851. 157562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 390e6"
142264. 185653
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 305c8"
184733. 220367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 385e7"
187817. 235498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 380f5"
232370. 276380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 313f4"
248926. 279011
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 425c2"
275910. 279011
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="cosmid 315g5"
<1..1932
/note="HS358B7"
complement(1294..1450)
/note="MIR; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/MIR"
1833. 39308
/note="HS316G12"
2453. 2501
/note="MER5A; RepeatMasker predicted
23/10/2000"
/rpt_family="DNA/MER1_type"
2517. 2817
/note="AluJo; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
2819. 2926
/note="MER3A; RepeatMasker predicted
23/10/2000"
/rpt_family="DNA/MER1_type"
2839. 4637
/note="CpG island; BAIAP3"
/evidence=not_experimental
3444. 3502

```

```
/note="GC-rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Low_complexity"
3559..19342
/ene="BAIAP3"
/ene="BAP3"
4319..4339
/note="GC-rich; RepeatMasker predicted
23/10/2000"
/rpt_family="Low_complexity"
9399..9506,10825..10884,11032..11098,11197..11392,
11725..11769,11889..11966,12084..12186,12430..12537,
12622..12717,12816..12919,13296..13416,13925..14049,
14130..14220,14372..14428,14504..14601,14701..14752,
14872..15002,15152..15275,15650..15793,15889..15981,
16073..16264,16345..16413,16508..16587,16750..16858,
17201..17272,17448..17528,17604..17724,17797..17920,
17999..18204,18282..19342)
/ene="BAIAP3"
/note="Similar to Unigene cluster Hs.101516"
Join(4565..4659,8443..8583,8805..8892,9067..9147,
9399..9506,10825..10884,11032..11098,11197..11392,
11725..11769,11889..11966,12084..12186,12430..12537,
12622..12717,12816..12919,13296..13416,13925..14049,
14130..14220,14372..14428,14504..14601,14701..14752,
14872..15002,15152..15275,15650..15793,15889..15981,
16073..16264,16345..16413,16508..16587,16750..16858,
17201..17272,17448..17528,17604..17724,17797..17920,
17999..18204,18282..18385)
/ene="BAIAP3"
/note="PubMed:9872452"
/codon_start=1
/product="BAI associated protein 3"
/protein_id="AAK61275.1"
/db_xref="GI:14336746"
/db_xref="MIM:604009"
/db_xref="SPTREMBL:O94812"
/translation="MRPRAFAAGPPDHLGTAIGAGATWRSRSPAMSTLLDIKS
SVLRQVQCPFRRTQDQGSADQEPATGAWKPGDGVFEFFAHMRLMLKKGGRQ
GLPCLEVRSGSPAPPEVDFSLGRALAPEEVEALYLYRAGTGPQDQV
DEALLSYLOQVFTSLEHTAERKAKATVALKVSVMKLNKAKDNGPQSDP
YCMGLIPASTDAPRAQKORFGRKSGRGLPAKCIQVTEKSTLNPVKEH
FLPEEDYSTDLHDIDWDHDDVSLVACRKLNEVIGLKGRTFKIVKSARANGT
AGPTEDHDDFLGCLNIPREVPAQVDRFKLEPRSSASRVQGHCHLVKLITTDQ
TAMSGRSGFLSHLLSHLLSHLSAEPSNWSRGELSTPAATILCLHGAQNL
PLQALVHWQVSSRHQCTLDYSYLLGLEDMQAHWEAPSLPQEESEADLSAF
SEFGLRLQRDLYFPATNSTAVHRLLELLKGLQLFQPSFCEPSELSNMDIAA
ALKRGNREWDRLNDKSPREPQORPLGLVLADAVDDLOFCYSVYASLFHILN
VDVTLFRQLERLYAEANVLTELSPKMTLEVASGLPELYLTLDLQRFWDSPGR
DSRLALAGIHAPFLFPAVKLWFQVLQDQAKLQAGVDMDTLEPVDASRRHSSAATA
GLCSLHQLVRLAWPQAPQAGTQGDVCEATLFTYELLKKYDTPGAAAGA
VSEALCVLNNVLRKAAGALKGLWEPATGEPGLPRLPCLSTQALDDDLQEA
HTVAHLSKMWGDIKRVQHSISPDSTONDEAVPLMKYLDKELALINSLVGNL
GSYLEWELLLOALQALGANRDVADYFSRFTLEALVSFFHAEGQPLESLRD
SVYELKELBLKCSRECEIOFYLDLKLQRTLEONREGLSVRECHVEAAQRLAVE
VLRADLPLDANGSDPFIYELGPPHFLPLVRSQRTQVKTRTLHYDELFYSVP
AEACRRRAACVLTFTYMDHLSLSTNDFAEAGLGLGVTVGVARPQVGGGARAGQPTLH
LCRRPRAACVLTFTYMDHLSLSTNDFAEAGLGLGVTVGVARPQVGGGARAGQPTLH
5744..5876
/note="MIR; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/MIR"
7123..7134
/note="AluSc; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
7135..7385
/note="AluSg; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
7386..7478
/note="AluSc; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
7572..7592
/note="(CTG)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
9646..9787
/note="(CA)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
9799..9977
/note="(CA)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
10020..10194
/note="(CA)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
10239..10411
/note="(CACC)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
10423..10514
Query Match 86.3%; Score 16.4; DB 9; Length 279011;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGGATCATCATGTGCGC 19
||||||| |||||||
DB 23876 GGGATCATCATGTGCGC 23893
||||||| |||||||

RESULT 9
AP001517/c
LOCUS 294250 bp DNA linear BCT 10-JAN-2001
DEFINITION Bacillus halodurans genomic DNA, section 11/14.
ACCESSION AP001517 BA000004
VERSION AP001517.1 GI:10175500
KEYWORDS
SOURCE Bacillus halodurans DNA.
ORGANISM Bacillus halodurans
REFERENCE 1 (sites)
AUTHORS Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.
TAKAMI, H., NAKASONE, K., HIRAMA, C., TAKAKI, Y., MASUI, N., FUJI, F.,
NAKAMURA, Y. and INOUE, A.
TITLE An improved physical and genetic map of the genome of alkaliphilic
Bacillus sp. C-125
JOURNAL Extremophiles 3 (1), 21-28 (1999)
MEDLINE 99184645
PUBMED 10086841
REFERENCE 2 (sites)
AUTHORS Takami, H. and Horikoshi, K.
TITLE Reidentification of facultatively alkaliphilic Bacillus sp. C-125
to Bacillus halodurans
JOURNAL Biosci. Biotechnol. Biochem. 63, 943-945 (1999)
REFERENCE 3 (sites)
AUTHORS Takami, H., Nakasone, K., Ogasawara, N., HIRAMA, C., NAKAMURA, Y.,
MASUI, N., FUJI, F., TAKAKI, Y., INOUE, A. and HORIKOSHI, K.
TITLE Sequencing of three lambda clones from the genome of alkaliphilic
Bacillus sp. strain C-125
JOURNAL Extremophiles 3 (1), 29-34 (1999)
MEDLINE 99184646
PUBMED 10086842
REFERENCE 4 (sites)
AUTHORS Takami, H., Takaki, Y., Nakasone, K., HIRAMA, C., INOUE, A. and
HORIKOSHI, K.
TITLE Sequence analysis of a 32-kb region including the major ribosomal
protein gene clusters from alkaliphilic Bacillus sp. strain C-125
Biosci. Biotechnol. Biochem. 63 (2), 452-455 (1999)
JOURNAL 99209008
MEDLINE 10192928
PUBMED 10192928
REFERENCE 5 (sites)
```

AUTHORS Takami, H., Takaki, Y., Nakasone, K., Sakiyama, T., Maeno, G.,  
 TITLE Genetic analysis of the chromosome of alkaliphilic *Bacillus*  
 halodurans C-125  
 JOURNAL Extremophiles 3 (3), 227-233 (1999)  
 MEDLINE 99411980  
 PUBMED 10484179  
 REFERENCE 6 (sites)  
 AUTHORS Takami, H.  
 TITLE Genome analysis of facultatively alkaliphilic *Bacillus* halodurans  
 C-125  
 JOURNAL (in) Extremophiles in deep-sea environments (Ed.);  
 HORIKOSHI, K. TSUJII;  
 : 249-284; Springer-Verlag (1999)  
 REFERENCE 7 (sites)  
 AUTHORS Takami, H., Masui, N., Nakasone, K. and Horikoshi, K.  
 TITLE Replication origin region of the chromosome of alkaliphilic  
*Bacillus* halodurans C-125  
 JOURNAL Biosci. Biotechnol. Biochem. 63 (6), 1134-1137 (1999)  
 MEDLINE 99356711  
 PUBMED 10427704  
 REFERENCE 8 (sites)  
 AUTHORS Takami, H. and Horikoshi, K.  
 TITLE Analysis of the genome of an alkaliphilic *Bacillus* strain from an  
 industrial point of view  
 JOURNAL Extremophiles 4 (2), 99-108 (2000)  
 MEDLINE 20263314  
 PUBMED 10805564  
 REFERENCE 9 (sites)  
 AUTHORS Nakasone, K., Masui, N., Takaki, Y., Sasaki, R., Maeno, G., Sakiyama, T.,  
 Hirama, C., Fujii, F. and Takami, H.  
 TITLE Characterization and comparative study of the *rrn* operons of  
 alkaliphilic *Bacillus* halodurans C-125  
 JOURNAL Extremophiles 4 (4), 209-214. (2000)  
 MEDLINE 20426005  
 PUBMED 10972189  
 REFERENCE 10 (sites)  
 AUTHORS Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N.,  
 Fujii, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and  
 Horikoshi, K.  
 TITLE Complete genome sequence of the alkaliphilic bacterium *Bacillus*  
 halodurans and genomic sequence comparison with *Bacillus* subtilis  
 Nucleic Acids Res. 28 (21), 4317-4331 (2000)  
 MEDLINE 20512582  
 PUBMED 11058132  
 REFERENCE 11 (bases 1 to 294250)  
 AUTHORS Takami, H. and Takaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-MAR-2000) Hideto Takami, Japan Marine Science and  
 Technology Center, Deep-sea Microorganisms Research Group; 2-15  
 Natsushima, Yokosuka, Kanagawa 237-0061, Japan  
 (E-mail: takami@jamstec.go.jp)  
 URL: <http://www.jamstec.go.jp/jamstec-e/bio/DEEPSTAR/Research.html>,  
 Tel: 81-468-67-3895, Fax: 81-468-66-6364)  
 FEATURES  
 source location/Qualifiers  
 1. 294250  
 /organism="Bacillus halodurans"  
 /db\_xref="taxon:86665"  
 /note="alkaliphile"  
 236. 823  
 /gene="BH2879"  
 236. 823  
 /gene="BH2879"  
 /codon\_start=1  
 /transl\_table=11  
 /product="spore cortex-lytic enzyme"  
 /protein\_id="BAB06598.1"  
 /db\_xref="GI:10175501"  
 /translation="MMKVFLLFSCIILSFLFTNKAASHKHQVQSGDTLYLLEQY  
 GVPEAKNEKSSNTIYRGELTIPARTASERDILARLVHAEAGEPYAGVAVA  
 VVYLNRVDHPSFDNDVINEVNSGYAPSPVONGIQRPADEAKRANDEALQFRG  
 OGNGLIFYNPKTATSPYVSTROQTIVIGNHFAK"  
 955. 1605

CDS

/gene="BH2880"  
 955. 1605  
 /gene="BH2880"  
 /note="BH2880  
 unknown"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB06599.1"  
 /db\_xref="GI:10175502"

gene

/gene="BH2881"  
 1694. 2977  
 /gene="BH2881"  
 1694. 2977  
 /gene="BH2881"  
 /note="BH2881  
 unknown conserved protein in others"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BAB06600.1"  
 /db\_xref="GI:10175503"

CDS

/gene="BH2882"  
 complement(3074. 4309)  
 /gene="BH2882"  
 complement(3074. 4309)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06601.1"  
 /db\_xref="GI:10175504"

gene

/gene="BH2883"  
 complement(4347. 5285)  
 /gene="BH2883"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06602.1"  
 /db\_xref="GI:10175505"

CDS

/gene="BH2884"  
 complement(4347. 5285)  
 /gene="BH2884"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06603.1"  
 /db\_xref="GI:10175506"

gene

/gene="BH2885"  
 complement(4347. 5285)  
 /gene="BH2885"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06604.1"  
 /db\_xref="GI:10175507"

CDS

/gene="BH2886"  
 complement(4347. 5285)  
 /gene="BH2886"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06605.1"  
 /db\_xref="GI:10175508"

gene

/gene="BH2887"  
 complement(4347. 5285)  
 /gene="BH2887"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06606.1"  
 /db\_xref="GI:10175509"

CDS

/gene="BH2888"  
 complement(4347. 5285)  
 /gene="BH2888"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06607.1"  
 /db\_xref="GI:10175510"

gene

/gene="BH2889"  
 complement(4347. 5285)  
 /gene="BH2889"  
 complement(4347. 5285)  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-oxoacyl-(acyl-carrier protein) synthase"  
 /protein\_id="BAB06608.1"  
 /db\_xref="GI:10175511"

/translation="MSNQRDFDTLMFGRPPSTDAEENPQESTSGSTVETKTE  
 GETKQETNAPALPLTFEQMEHIFRLAQSLGPAKLSLSPVVKTIQOLISQKKD"  
 complement(5903..6157)  
 /gene="BH2885"  
 complement(5903..6157)  
 /gene="BH2885"  
 /note="BH2885"  
 unknown  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="BA806604.1"  
 /db\_xref="GI:10175507"  
 /translation="MNVTYHGPIYRYGYIGTGERYPAYLPLFAGLAVGLPLFGKPK  
 YGYGPGAYGSPYGNFPGPNYGNFPGMYPGPPPGYYK"  
 complement(6226..6555)  
 /gene="BH2886"

Query Match 86.3%; Score 16.4; DB 1; Length 294250;  
 Best Local Similarity 94.4%; Pred. No. 3e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGGGGATCATGTCGCC 18  
 ||||| ||||| ||||| |||||  
 Db 279338 CGGGGATCATGTCGCC 279321

RESULT 10  
 HSDJ989D7  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP5-989D7 on chromosome 20. Contains  
 the 3' end of a novel gene, STSS and GSSs, complete sequence.  
 ACCESSION  
 VERSION  
 ALL121911  
 KEYWORDS  
 HTG  
 SOURCE  
 human.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Griffiths, C.  
 Direct Submission  
 Submitted (02-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonesrequest@sanger.ac.uk  
 On Aug 22, 2000 this sequence version replaced gi:9714378.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 20, constructed by the Sanger Centre Chromosome 20  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/RCP/Chr20  
 IMPORTANT: This sequence is not the entire insert of clone  
 RP5-989D7 it may be shorter because we sequence overlapping  
 sections only once, except for a 100 base overlap.  
 The true left end of clone RP5-989D7 is at 1 in this sequence. The  
 true left end of clone RP5-859D4 is at 119691 in this sequence. The  
 true right end of clone RP11-199014 is at 19977 in this sequence.  
 This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30);  
 an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. RP5-989D7 is from the

library RPCI-5 constructed by the group of Pieter de Jong. For  
 further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pCYPAC2.  
 FEATURES  
 source  
 Location/Qualifiers  
 1..119790  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /clone="RP5-989D7"  
 /clone\_lib="RPCI-5"  
 1..161  
 /gene="dj989D7.1"  
 <1..161  
 /gene="dj989D7.1"  
 /product="dj989D7.1 (novel transcript)"  
 /note="Continued from BA199014.1 in Em:AL162504  
 match: CDNAS: Em:AY007089"  
 /evidence="not\_experimental"  
 148..153  
 /gene="dj989D7.1"  
 945..3332  
 /note="TIGER1 repeat: matches 1..2418 of consensus"  
 4130..4386  
 /note="match: GSS: Em:AQ380734"  
 4294..4326  
 /note="L1MB8 repeat: matches 5834..5865 of consensus"  
 7201..8108  
 /note="L1M4 repeat: matches 4784..5705 of consensus"  
 complement(8165..8633)  
 /note="match: GSS: Em:AQ277139"  
 8629..10044  
 /note="L1M4 repeat: matches -260..1191 of consensus"  
 10045..10190  
 /note="73 copies 2 mer tt 57% conserved"  
 10382..10435  
 /note="27 copies 2 mer tt 74% conserved"  
 11843..12150  
 /note="Alusx repeat: matches 1..308 of consensus"  
 12350..12538  
 /note="MIR repeat: matches 60..249 of consensus"  
 12659..13160  
 /note="MLT1J repeat: matches 13..513 of consensus"  
 14687..14768  
 /note="HAL1 repeat: matches 271..360 of consensus"  
 15106..15438  
 /note="MER2 repeat: matches 1..345 of consensus"  
 15368..15768  
 /note="match: GSS: Em:AQ772240"  
 15375..15560  
 /note="match: GSS: Em:AQ012918"  
 15807..15507  
 /note="L1MB2 repeat: matches 5469..6165 of consensus"  
 16561..17220  
 /note="L1MB2 repeat: matches 4871..5474 of consensus"  
 17218..17509  
 /note="L1MB8 repeat: matches 5848..6168 of consensus"  
 17381..17733  
 /note="MER8 repeat: matches 87..239 of consensus"  
 17734..17987  
 /note="L1MB1 repeat: matches 5782..6043 of consensus"  
 18038..18183  
 /note="L1MA10 repeat: matches 6174..6316 of consensus"  
 18188..18525  
 /note="L1MB7 repeat: matches 2409..2740 of consensus"  
 18526..18878  
 /note="THE1A repeat: matches 1..354 of consensus"  
 18879..22631  
 /note="L1MB7 repeat: matches 2740..6162 of consensus"  
 23729..24007  
 /note="L1MB1 repeat: matches 5598..5892 of consensus"  
 24031..24127  
 /note="L1MA10 repeat: matches 5993..6095 of consensus"

```

repeat_region 24128..24439
/note="MER7A repeat: matches 1. .345 of consensus"
repeat_region 24440..24524
/note="L1MA10 repeat: matches 6095. .6317 of consensus"
repeat_region 25218..27841
/note="L1MCI repeat: matches 3550. .6326 of consensus"
misc_feature complement(27586..28074)
/note="match: GSS: Em:AQ605230"
repeat_region 28283..28491
/note="L1R16A repeat: matches 148. .378 of consensus"
repeat_region 28592..28729
/note="MIR repeat: matches 26. .167 of consensus"
repeat_region 29416..29585
/note="MER5B repeat: matches 1. .177 of consensus"
repeat_region 30385..30489
/note="L1R16C repeat: matches 81. .221 of consensus"
repeat_region 32510..33153
/note="MER51B repeat: matches 1. .617 of consensus"
repeat_region 33154..33323
/note="L1R16C repeat: matches 221. .385 of consensus"
misc_feature 35252..35712
/note="match: GSS: Em:AQ700733"
misc_feature 35693..36037
/note="match: STS: Em:L16414"
repeat_region 35731..35802
/note="MTL1J repeat: matches 124. .196 of consensus"
misc_feature 35934..36427
/note="match: GSS: Em:AQ429711"
repeat_region 35967..36020
/note="9 copies 6 mer tagata 75% conserved"
repeat_region 35968..36019
/note="13 copies 4 mer agat 100% conserved"
repeat_region 36212..36570
/note="L1R16A repeat: matches 57. .433 of consensus"
misc_feature 38224..38413
/note="match: GSS: Em:A2235965"
repeat_region 38910..39037
/note="MIR repeat: matches 74. .225 of consensus"
repeat_region 39063..39103
/note="L2 repeat: matches 2705. .2745 of consensus"
repeat_region 39217..39290
/note="L2 repeat: matches 2410. .2484 of consensus"
repeat_region 39291..39673
/note="MTL1A1 repeat: matches 4. .365 of consensus"
repeat_region 39678..39933
/note="L2 repeat: matches 2159. .2419 of consensus"
repeat_region 39960..40082
/note="L2 repeat: matches 1978. .2096 of consensus"
repeat_region 40421..41176
/note="L1P16 repeat: matches 3166. .3932 of consensus"
repeat_region 41179..43371
/note="L1P16 repeat: matches 3923. .6157 of consensus"
repeat_region 43715..43783
/note="MIR repeat: matches 14. .80 of consensus"
repeat_region 45011..45066
/note="14 copies 4 mer tata 80% conserved"
repeat_region 45244..45466
/note="L2 repeat: matches 2168. .2393 of consensus"
repeat_region 45535..45726
/note="L2 repeat: matches 2557. .2746 of consensus"
repeat_region 46043..46146
/note="L2 repeat: matches 2602. .2710 of consensus"
repeat_region 46287..46523

```

```

/note="L2 repeat: matches 2022. .2266 of consensus"
46738..46912
/note="L2 repeat: matches 2541. .2710 of consensus"
47299..47342
/note="MIR repeat: matches 97. .139 of consensus"
47331..47479
/note="MIR repeat: matches 99. .256 of consensus"
complement(47625..48021)
/note="match: GSS: Em:AQ818871"

Query Match      84.2%; Score 16; DB 9; Length 119790;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGGGGATCAGTGGTG 16
    |||||
Db 101041 CGGGGATCAGTGGTG 101056

RESULT 11
AC026171/c
LOCUS AC026171 160819 bp DNA linear HTG 29-MAY-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-166F3 map 3p, WORKING DRAFT
SEQUENCE, 38 unordered pieces.
ACCESSION AC026171
VERSION AC026171.2 GI:8101140
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 160819)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160819)
AUTHORS Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,
Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J., Li, L.,
Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, Y.,
Luo, J., Niu, Y., Qi, O., Qi, X., Song, S., Sun, M., Sun, W., Sun, Y.,
Tao, R., Wang, H., Wang, J., Wang, J., Wang, L., Wang, L., Wang, R.,
Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,
Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,
Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,
Zhang, Z., Zhu, B., Yu, J. and Yang, H.
Chromosome 3p genomic sequence.
Unpublished
2 (bases 1 to 160819)
AUTHORS Kang, N., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,
Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,
Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,
Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.
and Yang, H.
Direct Submission
Submitted (21-MAR-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On May 29, 2000 this sequence version replaced gi:7271988.
COMMENT
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgsc.igtp.ac.cn
http://www.genomics.org.cn
Contact:hgsc@igtp.ac.cn
-----Project Information
Center project name:1% project
Center clone name: RP11-166F3
-----Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 127362 bases at least Q40
Consensus quality: 138470 bases at least Q30
Consensus quality: 147433 bases at least Q20

```

Insert size: 125850; sum-of-contigs  
Quality coverage: 3.22x in Q20 bases; sum-of-contigs

1	1175: contig of 1175 bp in length
1176	1275: gap of unknown length
1276	2379: contig of 1104 bp in length
2380	2479: gap of unknown length
2480	3439: contig of 1884 bp in length
4364	4463: gap of unknown length
4464	6504: contig of 2041 bp in length
6505	6604: gap of unknown length
6605	8243: contig of 1638 bp in length
8243	8342: gap of unknown length
8343	10668: contig of 2326 bp in length
10669	10768: gap of unknown length
10769	12767: contig of 1999 bp in length
12768	12867: gap of unknown length
12868	15235: contig of 2368 bp in length
15236	15335: gap of unknown length
15336	17283: contig of 1948 bp in length
17284	17383: gap of unknown length
17384	19642: contig of 2259 bp in length
19643	19742: gap of unknown length
19743	2252: contig of 2780 bp in length
22523	22622: gap of unknown length
22623	25720: contig of 3098 bp in length
25721	25820: gap of unknown length
25821	29319: contig of 3499 bp in length
29320	29419: gap of unknown length
29420	31356: contig of 1937 bp in length
31357	31456: gap of unknown length
31457	33345: contig of 1889 bp in length
33346	33445: gap of unknown length
33446	35867: contig of 2422 bp in length
35868	35967: gap of unknown length
35968	38450: contig of 2483 bp in length
38451	38550: gap of unknown length
38551	41348: contig of 2798 bp in length
41349	41448: gap of unknown length
41449	45413: contig of 3965 bp in length
45414	45513: gap of unknown length
45514	49633: contig of 4120 bp in length
49634	49733: gap of unknown length
49734	53847: contig of 4114 bp in length
53848	53947: gap of unknown length
53948	58507: contig of 4560 bp in length
58508	58607: gap of unknown length
58608	62353: contig of 3746 bp in length
62354	62453: gap of unknown length
62454	66817: contig of 4304 bp in length
66818	66917: gap of unknown length
66918	71210: contig of 4293 bp in length
71211	71310: gap of unknown length
71311	75769: contig of 4459 bp in length
75770	75869: gap of unknown length
75870	82021: contig of 6152 bp in length
82022	82121: gap of unknown length
82122	87770: contig of 5649 bp in length
87771	87870: gap of unknown length
87871	93303: contig of 5433 bp in length
93304	93403: gap of unknown length
93404	98401: contig of 4998 bp in length
98402	98501: gap of unknown length
98502	104179: contig of 5678 bp in length
104180	104279: gap of unknown length

```

* 104280 111860: contig of 7581 bp in length
* 111861 111960: gap of unknown length
* 111961 119359: contig of 7399 bp in length
* 119360 119459: gap of unknown length
* 119460 129240: contig of 9781 bp in length
* 129241 129340: gap of unknown length
* 129341 135533: contig of 6195 bp in length
* 135533 142841: gap of unknown length
* 142841 142841: contig of 7206 bp in length
* 142842 150041: gap of unknown length
* 150042 150141: contig of 7100 bp in length
* 150142 160819: contig of 10678 bp in length.
FEATURES
    source
        1. 160819
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /map="3p"
            /clone="RP11-166f3"
            1. 11175
                /note="assembly_name:Contig10"
                1276..2379
                /note="assembly_name:Contig14"
                2480..4363
                /note="assembly_name:Contig16"
                4464..6504
                /note="assembly_name:Contig17"
                6605..8242
                /note="assembly_name:Contig18"
                8343..10668
                /note="assembly_name:Contig19"
                10769..12767
                /note="assembly_name:Contig21"
                12868..15235
                /note="assembly_name:Contig22"
                15336..17283
                /note="assembly_name:Contig23"
                17384..19642
                /note="assembly_name:Contig24"
                19743..22522
                /note="assembly_name:Contig25"
                22623..25720
                /note="assembly_name:Contig26"
                25821..29319
                /note="assembly_name:Contig27"
                29420..31356
                /note="assembly_name:Contig28"
                clone_end:77
                vector_slide:left"
                31457..33345
                /note="assembly_name:Contig29"
                33446..35867
                /note="assembly_name:Contig30"
                35968..38450
                /note="assembly_name:Contig31"
                38551..41348
                /note="assembly_name:Contig32"
                41449..45413
                /note="assembly_name:Contig33"
                45514..49633
                /note="assembly_name:Contig34"
                49734..53847
                /note="assembly_name:Contig35"
                53948..58507
                /note="assembly_name:Contig36"
                58608..62353
                /note="assembly_name:Contig37"
                62454..66817
                /note="assembly_name:Contig38"
                66918..71210
                /note="assembly_name:Contig39"
                71311..75769

```

```
misc_feature /note="assembly_name:Contig40"
75970..82021

Query Match      84.2%; Score 16; DB 2; Length 160819;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GGATCACATGGTGCCG 19
|||||
Db 38906 GGATCACATGGTGCCG 38991

RESULT 12
AR206865/c
LOCUS AR206865 504 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 3 from patent US 6372459.
ACCESSION AR206865
VERSION AR206865.1 GI:21505596
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 504)
AUTHORS Versek,S., Kula,M.-R., Bommarious,A. and Drauz,K.
TITLE N-acetyl amino acid racemase
JOURNAL Patent: US 6372459-A 3 16-APR-2002;
FEATURES Location/Qualifiers
source 1..504
/organism="unknown"
BASE COUNT 77 a 163 c 185 g 79 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 6; Length 504;
Best Local Similarity 89.5%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGGTGCCG 19
|||||
Db 220 CGGGGATCACATGGTGCCG 202

RESULT 13
AX136894/c
LOCUS AX136894 504 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 3 from Patent EP1074628.
ACCESSION AX136894
VERSION AX136894.1 GI:14273243
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 504)
AUTHORS Versek,S., Kula,M.-R., Bommarious,A. and Drauz,K.
TITLE N-acetyl amino acid racemase
JOURNAL Patent: EP 1074628-A 3 07-FEB-2001;
FEATURES Location/Qualifiers
source 1..504
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Gensonde"
BASE COUNT 77 a 163 c 185 g 79 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 6; Length 504;
Best Local Similarity 89.5%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGGTGCCG 19
|||||
Db 220 CGGGGATCACATGGTGCCG 202

RESULT 14
BD006820/c
LOCUS BD006820 504 bp DNA linear PAT 31-JAN-2002
DEFINITION N-Acetyl amino acid racemase, gene encoding the same, plasmid, vector and microorganism having the gene, primer and probe for the gene and utilization of the racemase.
ACCESSION BD006820
VERSION BD006820.1 GI:18635191
KEYWORDS JP 2001046088-A/2.
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 504)
AUTHORS Ferzeck,S., Kula,M.-R., Bommarious,A. and Dorautz,C.
TITLE N-Acetyl amino acid racemase, gene encoding the same, plasmid, vector and microorganism having the gene, primer and probe for the gene and utilization of the racemase
JOURNAL Patent: JP 2001046088-A 2 20-FEB-2001;
COMMENT DEGUSSA HUELS AG
OS Artificial Sequence
PN JP 2001046088-A/2
PD 20-FEB-2001
PF 24-JUL-2000 JP 2000222928
PI 27-JUL-1999 DE 19935268.2
PI STEPHEN FERZECK,MARIA REGINA KULA,ANDREAS BOMMARIOUS, PI CURLHEINTZ DORAUTZ
PC C12N15/09,C12N1/21,C12N9/90,C12P13/04/(C12N15/09,C12R1:04),
PC (C12N1/21,C12R1:04),(C12N9/90,C12R1:04),(C12P13/04,C12R1:04),
PC C12N15/00,
PC (C12N15/00,C12R1:04)
CC

FEATURES
source 1..504
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 77 a 163 c 185 g 79 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 6; Length 504;
Best Local Similarity 89.5%; Pred. No. 4.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGGTGCCG 19
|||||
Db 220 CGGGGATCACATGGTGCCG 202

RESULT 15
AR206864/c
LOCUS AR206864 1107 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6372459.
ACCESSION AR206864
VERSION AR206864.1 GI:21505594
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1107)
AUTHORS Versek,S., Kula,M.-R., Bommarious,A. and Drauz,K.
TITLE N-acetyl amino acid racemase
JOURNAL Patent: US 6372459-A 1 16-APR-2002;
FEATURES Location/Qualifiers
source 1..1107
/organism="unknown"
BASE COUNT 164 a 369 c 404 g 170 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 6; Length 1107;
Best Local Similarity 89.5%; Pred. No. 5.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 CCGGGATCACATGGTCCG 19  
||||||| |||||  
Db 220 CCGGGATCACATGGTCCG 202

Search completed: July 8, 2003, 03:34:32  
Job time : 214.093 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 119.376 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-8  
Perfect score: 19  
Sequence: 1. cgggatacattggtgcg 19

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	100.0	19	AAZ28814	Rat membrane metal
2	19	100.0	21	AAZ28816	Rat membrane metal
3	19	100.0	2765	AAZ28810	Rat membrane metal
4	16.4	86.3	410	AAI90923	Human polynucleoti
5	16.4	86.3	3939	AAI90923	Human polynucleoti
6	16.4	86.3	3939	AAI90923	DNA encoding novel
7	16.4	86.3	4829	AAI90923	DNA encoding novel
8	15.8	83.2	504	AAF61121	Human immune/haema
9	15.8	83.2	1107	AAF61120	A. orientalis subs

C 10	15.8	83.2	1107	24	AAI47264	A. orientalis N-ace
C 11	15.8	83.2	1107	24	AAI45919	Amycolatopsis orie
C 12	15.4	81.1	592	23	ABL28855	Drosophila melanog
C 13	15.4	81.1	970	22	AAI26105	Human cDNA encodin
C 14	15.4	81.1	2286	24	AAI28130	Soluble secreted e
C 15	15.4	81.1	4183	23	ABL28854	Drosophila melanog
C 16	15.4	81.1	16748	22	AAI65193	Human immune/haema
C 17	15	78.9	1324	23	ABL14849	Drosophila melanog
C 18	15	78.9	3474	23	ABL14848	Drosophila melanog
C 19	14.8	77.9	269	24	ABN25649	Human OREX polynuc
C 20	14.8	77.9	499	22	AAI93666	Human polynucleoti
C 21	14.8	77.9	831	24	ABL87041	Human ovarian canc
C 22	14.8	77.9	880	22	AAI94345	Human neuroblastom
C 23	14.8	77.9	898	22	AAI94317	Human neuroblastom
C 24	14.8	77.9	1347	17	AAI41698	Drosophila melanog
C 25	14.8	77.9	1347	17	AAI41699	Onchocerca volvulu
C 26	14.8	77.9	1347	17	AAI41700	Brugia pahangi bet
C 27	14.8	77.9	1371	20	AAI90953	Redesigned Aspergi
C 28	14.8	77.9	1667	24	ABA95078	Mouse IDPm encodin
C 29	14.8	77.9	1679	18	AAI51370	Elmeria lactate de
C 30	14.8	77.9	1912	17	AAI17715	Heat resistant mal
C 31	14.8	77.9	1977	17	AAI16307	Rice HY4 cDNA part
C 32	14.8	77.9	2220	24	ABO73246	Lolium perenne lpi
C 33	14.8	77.9	2331	22	AAK66148	Human immune/haema
C 34	14.8	77.9	2588	24	ABO73245	Lolium perenne lpi
C 35	14.8	77.9	2688	23	ABL13205	Drosophila melanog
C 36	14.8	77.9	2925	21	AAI63763	cDNA encoding neut
C 37	14.8	77.9	3424	24	ABI99398	Mouse ischaemic co
C 38	14.8	77.9	4226	22	ABA21324	Human nervous syst
C 39	14.8	77.9	4600	20	AAI03043	Human IL-1ra BAC c
C 40	14.8	77.9	4968	23	ABL13204	Drosophila melanog
C 41	14.8	77.9	6210	20	AAI02986	Human IL-1ra BAC c
C 42	14.8	77.9	20987	22	AAI06039	Human reproductive
C 43	14.8	77.9	20987	23	ABL98604	Human testicular a
C 44	14.8	77.9	24533	22	AAI27689	DNA encoding novel
C 45	14.4	75.8	97	22	ABA70793	Human foetal liver

ALIGNMENTS

RESULT 1	
AAZ28814	
ID	AAZ28814 standard; DNA; 19 BP.
AC	AAZ28814;
DT	01-FEB-2000 (first entry)
DE	Rat membrane metalloprotease NEPII gene probe #4.
DE	Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;
DE	neuron; hormone; peptide messenger; inhibitor; detection; disorder;
DE	cardiovascular disease; neurodegenerative disease; growth disorder;
DE	hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.
OS	Synthetic.
OS	Rattus rattus.
PN	FR2777291-A1.
PD	15-OCT-1999.
PF	08-APR-1998; 98FR-0004389.
PR	08-APR-1998; 98FR-0004389.
PA	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PI	Quimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;
PI	Schwartz JC;
PI	WPI; 1999-593429/51.

XX PT New membrane metalloprotease NEP II, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 3; Page 20; 29pp; French.  
 XX  
 CC Sequences AA228811-228827 represent probes for detecting the rat  
 CC membrane metalloprotease designated neprilysine II (NEPII) gene  
 CC (AA228810). NEPII is involved in (in)activation of neuronal and hormonal  
 CC peptide messengers. NEPII is used to screen for specific substrates (used  
 CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
 CC used to detect NEPII or for treatment of disorders related to peptidergic  
 CC signalling in which NEPII is involved, e.g. cardiovascular or  
 CC neurodegenerative diseases; growth disorders of endocrine origin;  
 CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
 CC disorders.  
 XX  
 SQ Sequence 19 BP; 3 A; 5 C; 8 G; 3 T; 0 other;  
 Query Match 100.0%; Score 19; DB 20; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGGGATCACATGGTGCCG 19  
 DB 1 CGGGGATCACATGGTGCCG 19  
 AA228816/c  
 ID AA228816 standard; DNA; 21 BP.  
 XX  
 AC AA228816;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Rat membrane metalloprotease NEPII gene probe #6.  
 XX  
 KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
 KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
 KW cardiovascular disease; neurodegenerative disease; growth disorder;  
 KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
 XX  
 OS Synthetic.  
 OS Rattus rattus.  
 XX  
 PN FR2777291-A1.  
 XX  
 PD 15-OCT-1999.  
 XX  
 PF 08-APR-1998; 98FR-0004389.  
 XX  
 PR 08-APR-1998; 98FR-0004389.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
 PI Schwartz JC;  
 XX  
 DR WPI; 1999-593429/51.  
 XX  
 PT New membrane metalloprotease NEP II, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 3; Page 21; 29pp; French.  
 XX  
 CC Sequences AA228811-228827 represent probes for detecting the rat  
 CC membrane metalloprotease designated neprilysine II (NEPII) gene  
 CC (AA228810). NEPII is involved in (in)activation of neuronal and hormonal  
 CC peptide messengers. NEPII is used to screen for specific substrates (used  
 CC to detect NEPII in cells and tissues) or inhibitors, which can also be

CC used to detect NEPII or for treatment of disorders related to peptidergic  
 CC signalling in which NEPII is involved, e.g. cardiovascular or  
 CC neurodegenerative diseases; growth disorders of endocrine origin;  
 CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
 CC disorders.  
 XX  
 SQ Sequence 21 BP; 4 A; 8 C; 6 G; 3 T; 0 other;  
 Query Match 100.0%; Score 19; DB 20; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGGGATCACATGGTGCCG 19  
 DB 1 CGGGGATCACATGGTGCCG 1  
 AA228810/c  
 ID AA228810 standard; cDNA; 2765 BP.  
 XX  
 AC AA228810;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Rat membrane metalloprotease NEPII gene.  
 XX  
 KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
 KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
 KW cardiovascular disease; neurodegenerative disease; growth disorder;  
 KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
 XX  
 OS Rattus rattus.  
 XX  
 PN FR2777291-A1.  
 XX  
 PD 15-OCT-1999.  
 XX  
 PF 08-APR-1998; 98FR-0004389.  
 XX  
 PR 08-APR-1998; 98FR-0004389.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
 PI Schwartz JC;  
 XX  
 DR WPI; 1999-593429/51.  
 XX  
 PD P-PSDB; AA244177.  
 XX  
 PT New membrane metalloprotease NEP II, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 2; Page 12-16; 29pp; French.  
 XX  
 CC This sequence represents the gene for the rat membrane metalloprotease  
 CC designated neprilysine II (NEPII), which is involved in (in)activation  
 CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
 CC for specific substrates (used to detect NEPII in cells and tissues) or  
 CC inhibitors, which can also be used to detect NEPII or for treatment of  
 CC disorders related to peptidergic signalling in which NEPII is involved,  
 CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
 CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
 CC endocrine disorders.  
 XX  
 SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
 Query Match 100.0%; Score 19; DB 20; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 3;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGGGGATCACATGGTGCCG 19

Db 485 CGGGATCACATGGTGCCG 467  
|||||

## RESULT 4

AAI90923  
ID AAI90923 standard; cDNA; 410 BP.

XX AC AAI90923;

XX DT 06-NOV-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 10983.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
tissue growth factor; immunomodulatory; cancer; leukaemia;  
nervous system disorders; arthritis; inflammation; ss.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-514838/56.

XX P-PSDB: AAO10992.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
diagnosing and treating e.g. leukaemia, inflammation and immune  
disorders

XX PS Claim 1; SEQ ID NO 10983; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
cytokine, cell proliferation or cell differentiation or which may induce  
production of other cytokines in other cell populations. The  
polynucleotides and polypeptides are useful in gene therapy, vaccines or  
peptide therapy. The polypeptides have various cytokine-like activities,  
e.g. stem cell growth factor activity, haematopoiesis, regulating  
activity, tissue growth factor activity, immunomodulatory activity and  
activin/inhibin activity and may be useful in the diagnosis and/or  
treatment of cancer, leukaemia, nervous system disorders, arthritis and  
inflammation.

XX CC Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 410 BP; 101 A; 81 C; 136 G; 92 T; 0 other;

Query Match 86.3%; Score 16.4; DB 22; Length 410;

Best Local Similarity 94.4%; Pred. No. 60;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGCCG 19

|||||

Db 211 GGGGATCAGATGGTGCCG 228

## RESULT 5

AAS88996/c

ID AAS88996 standard; cDNA; 3939 BP.

XX

AC AAS88996;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #24800.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG24809.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity

XX PS Claim 1; SEQ ID NO 24800; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.

XX CC The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
diagnostic coding sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3939 BP; 933 A; 1072 C; 1040 G; 894 T; 0 other;

Query Match 86.3%; Score 16.4; DB 23; Length 3939;

Best Local Similarity 94.4%; Pred. No. 69;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGCCG 19

|||||

Db 3807 GGGGATCACATGGTGCCG 3790

## RESULT 6

AAS94171/c

ID AAS94171 standard; cDNA; 3939 BP.

XX AC AAS94171;

XX

DT 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #29975.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG29984.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX Claim 1; SEQ ID NO 29975; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 3939 BP; 933 A; 1072 C; 1040 G; 894 T; 0 other;  
Query Match 86.3%; Score 16.4; DB 23; Length 3939;  
Best Local Similarity 94.4%; Pred. No. 69;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GGGGATCACATGGTGCGG 19  
IIIIIIIIIIIIIIIIIIII  
Db 3807 GGGGATCACATGGTGCGG 3790  
RESULT 7  
AAK84366  
ID AAK84366 standard; DNA; 4829 BP.  
XX  
AC AAK84366;  
XX  
DT 07-NOV-2001 (first entry)  
XX

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39178.  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01354.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.

```
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0248527.
PR 08-NOV-2000; 2000US-0248528.
PR 08-NOV-2000; 2000US-0248532.
PR 08-NOV-2000; 2000US-0248609.
PR 08-NOV-2000; 2000US-0248610.
PR 08-NOV-2000; 2000US-0248611.
PR 08-NOV-2000; 2000US-0248613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.

PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis
XX
XX Disclosures; SEQ ID NO 39178; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 4829 BP; 922 A; 1179 C; 1681 G; 1047 T; 0 other;

Query Match 86.3%; Score 16.4; DB 22; Length 4829;
Best Local Similarity 94.4%; Pred. No. 70;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGATCAGATGGTGGCG 19
|||||||
DB 1006 GGGGATCAGATGGTGGCG 1023

RESULT 8
AAF61121/c
ID AAF61121 standard; DNA; 504 BP.
XX
AC AAF61121.
XX
XX 17-MAY-2001 (first entry)
XX
XX A. orientalis PCR generated N-acetyl amino acid racemase DNA.
XX
XX N-acetyl amino acid racemase; AAR; enantiomerically enriched amino acid;
KW enzyme-membrane reactor; N-acetyl-D-methionine; N-acetyl-L-methionine;
KW L-methionine; heavy metal dependency; ds.
XX
XX Anycolatopsis orientalis.
XX
XX EP1074628-A1.
XX
XX 07-FEB-2001.
XX
XX 25-JUL-2000; 2000EP-0115902.
XX
XX 27-JUL-1999; 99DE-1035268.
XX
XX (DEGS ) DEGUSSA-HUELS AG.
XX
XX Verseck S, Kula M, Bommarius A, Drauz K;
PI
XX
```

DR WPI; 2001-161182/17.  
 XX New N-acetyl amino acid racemase enzyme derived from Amycolatopsis  
 PT orientalis ssp. lurida, useful for producing enantiomerically enriched  
 PT amino acids  
 XX  
 PS Disclosure; Page 13; 23pp; German.  
 XX  
 CC This invention describes a novel N-acetyl amino acid racemase (AAR)  
 CC enzyme (I) derived from Amycolatopsis orientalis ssp. lurida (DSM 43134).  
 CC The invention also describes (1) a gene coding for (I); (2) a vector  
 CC containing the gene; (3) a microorganism containing the gene; (4) a  
 CC primer for the gene; and (5) a probe for the gene. (I) is useful for  
 CC producing enantiomerically enriched amino acids in an enzyme-membrane  
 CC reactor, e.g. by AAR-catalyzed conversion of N-acetyl-D-methionine to  
 CC L-methionine. (I) exhibits reduced heavy metal dependency compared with  
 CC the AAR of Amycolatopsis sp. TS-1-60 (Appl. Microbiol. Biotechnol., 42,  
 CC 853, 1995).  
 XX  
 SQ Sequence 504 BP; 77 A; 163 C; 185 G; 79 T; 0 other;  
 Query Match 83.2%; Score 15.8; DB 22; Length 504;  
 Best Local Similarity 89.5%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CGGGGATCAGATGGTGCCG 19  
 DB 220 CGGGGATCAGATGGTGCCG 202  
 RESULT 9  
 AAF61120/c  
 ID AAF61120 standard; DNA; 1107 BP.  
 XX  
 AC AAF61120;  
 XX  
 DT 17-MAY-2001 (first entry)  
 XX  
 DE A. orientalis subsp lurida N-acetyl amino acid racemase DNA.  
 XX  
 KW N-acetyl amino acid racemase; AAR; enantiomerically enriched amino acid;  
 KW enzyme-membrane reactor; N-acetyl-D-methionine; N-acetyl-L-methionine;  
 KW L-methionine; heavy metal dependency; ds.  
 XX  
 OS Amycolatopsis orientalis.  
 XX  
 PN EP1074628-A1.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 25-JUL-2000; 2000EP-0115902.  
 XX  
 PR 27-JUL-1999; 99DE-1035268.  
 XX  
 PA (DEGS ) DEGUSSA-HUELS AG.  
 XX  
 PI Verseck S, Kula M, Bommarius A, Drauz K;  
 XX  
 DR WPI; 2001-161182/17.  
 DR P-PSDB; AAB70760.  
 XX  
 CC New N-acetyl amino acid racemase enzyme derived from Amycolatopsis  
 PT orientalis ssp. lurida, useful for producing enantiomerically enriched  
 PT amino acids  
 XX  
 PS Disclosure; Page 10-11; 23pp; German.  
 XX  
 CC This invention describes a novel N-acetyl amino acid racemase (AAR)  
 CC enzyme (I) derived from Amycolatopsis orientalis ssp. lurida (DSM 43134).  
 CC The invention also describes (1) a gene coding for (I); (2) a vector  
 CC containing the gene; (3) a microorganism containing the gene; (4) a  
 CC primer for the gene; and (5) a probe for the gene. (I) is useful for

CC producing enantiomerically enriched amino acids in an enzyme-membrane  
 CC reactor, e.g. by AAR-catalyzed conversion of N-acetyl-D-methionine to  
 CC N-acetyl-L-methionine followed by acylase-catalyzed conversion to  
 CC L-methionine. (I) exhibits reduced heavy metal dependency compared with  
 CC the AAR of Amycolatopsis sp. TS-1-60 (Appl. Microbiol. Biotechnol., 42,  
 CC 853, 1995).  
 XX  
 SQ Sequence 1107 BP; 164 A; 369 C; 404 G; 170 T; 0 other;  
 Query Match 83.2%; Score 15.8; DB 22; Length 1107;  
 Best Local Similarity 89.5%; Pred. No. 1.3e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CGGGGATCAGATGGTGCCG 19  
 DB 220 CGGGGATCAGATGGTGCCG 202  
 RESULT 10  
 AAL47264/c  
 ID AAL47264 standard; DNA; 1107 BP.  
 XX  
 AC AAL47264;  
 XX  
 DT 05-SEP-2002 (first entry)  
 XX  
 DE A orientalis N-acetyl amino acid racemase coding sequence.  
 XX  
 KW N-acetyl amino acid racemase; enantiomer-enriched; N-protected;  
 KW bioactive; parenteral nutrition; gene; ds.  
 XX  
 OS Amycolatopsis orientalis.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1107  
 FT /\*tag= a  
 FT /product= "N-acetyl amino acid racemase"  
 FT /partial  
 FT /note= "No start codon"  
 XX  
 PN EP1197563-A1.  
 XX  
 PD 17-APR-2002.  
 XX  
 PF 11-OCT-2001; 2001EP-0124428.  
 XX  
 PR 11-OCT-2000; 2000DE-1050123.  
 XX  
 PA (DEGS ) DEGUSSA AG.  
 XX  
 PI Bommarius A, Verseck S, Drauz K;  
 XX  
 DR WPI; 2002-492097/53.  
 DR P-PSDB; AAO17910.  
 XX  
 PT Preparation of enantiomer-enriched amino acids comprising contacting  
 PT N-protected amino acid with N-acetyl amino acid racemase enzyme and  
 PT amino acid acylase enzyme  
 XX  
 PS Disclosure; Page 5-6; 12pp; German.  
 XX  
 CC The present invention relates to the preparation of enantiomer-enriched  
 CC amino acids and involves contacting an N-protected alpha-amino acid with  
 CC an N-acetyl amino acid racemase enzyme and an amino acid acylase enzyme.  
 CC The enantiomer-enriched amino acids are useful in the synthesis of  
 CC bioactive compounds and for parenteral nutrition. The present sequence is  
 CC the Amycolatopsis orientalis N-acetyl amino acid racemase coding  
 CC sequence.  
 XX  
 SQ Sequence 1107 BP; 164 A; 369 C; 404 G; 170 T; 0 other;  
 Query Match 83.2%; Score 15.8; DB 24; Length 1107;  
 Best Local Similarity 89.5%; Pred. No. 1.3e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCATGGTCCG 19  
 DE ||||| ||||| ||||| |||||  
 Db 220 CGGGGATCATGGTCCG 202

RESULT 11  
 AAL45919/C  
 ID AAL45919 standard; DNA; 1107 BP.  
 AC AAL45919;  
 DT 19-JUL-2002 (first entry)  
 DE Amycolatopsis orientalis N-acetylaminic acid racemase coding sequence.  
 KW N-acetylaminic acid racemase; AAR; racemisation; N-carbamoyl-amino acid;  
 KW enantiomerically enriched amino acid; chemical synthesis; enzyme;  
 KW parenteral nutrition; gene; ds.  
 XX OS Amycolatopsis orientalis.

XX Key Location/Qualifiers  
 FH CDS 1..1107  
 FT /\*tag= a  
 FT /product= "N-acetylaminic acid racemase"  
 FT /note= "no stop codon"  
 XX EP1199369-A2.  
 XX PD 24-APR-2002.  
 XX PF 04-OCT-2001; 2001EP-0123744.  
 XX PR 11-OCT-2000; 2000DE-1050124.  
 XX PA (DEGS ) DEGUSSA AG.  
 XX PI Bommarius A, Verseck S, Drauz K, Kula M;  
 XX WPI: 2002-364535/40.  
 XX DR P-PSDB; AA017352.

XX Racemizing N-carbamoyl-amino acids, useful in production of pure amino  
 PT acid enantiomers, comprises using N-acetylaminic acid racemase from  
 PT specific Amycolatopsis orientalis  
 XX Disclosure: Page 5-6; 8pp; German.  
 XX The present invention relates to the use of the N-acetylaminic acid  
 CC racemase (AAR) from Amycolatopsis orientalis subspecies lurida for  
 CC racemisation of N-carbamoyl-amino acids. The method is used in the  
 CC production of enantiomerically enriched amino acids, which are useful for  
 CC chemical synthesis or parenteral nutrition. The present sequence is the  
 CC A. orientalis AAR coding sequence.

XX SQ Sequence 1107 BP; 164 A; 369 C; 404 G; 170 T; 0 other;  
 Query Match 83.2%; Score 15.8; DB 24; Length 1107;  
 Best Local Similarity 89.3%; Pred. No. 1.3e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCATGGTCCG 19  
 DE ||||| ||||| ||||| |||||  
 Db 220 CGGGGATCATGGTCCG 202

RESULT 12  
 ABL28855  
 ID ABL28855 standard; DNA; 592 BP.  
 XX ABL28855;  
 AC

XX DT 26-MAR-2002 (first entry)  
 XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 38038.  
 XX KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX PD 27-SEP-2001.  
 XX PF 23-MAR-2001; 2001WO-US09231.  
 XX PR 23-MAR-2000; 2000US-191637P.  
 XX PR 11-JUL-2000; 2000US-0614150.  
 XX PA (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI: 2001-656860/75.  
 XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions  
 XX PS Claim 1; SEQ ID NO 38038; 21pp + Sequence Listing; English.  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 592 BP; 133 A; 196 C; 181 G; 82 T; 0 other;  
 Query Match 81.1%; Score 15.4; DB 23; Length 592;  
 Best Local Similarity 94.1%; Pred. No. 2e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGATCATGGTCCG 19  
 DE ||||| ||||| ||||| |||||  
 Db 503 GGGATCATGGTCCG 519

RESULT 13  
 AAS26105  
 ID AAS26105 standard; cDNA; 970 BP.  
 AC AAS26105;  
 XX 07-NOV-2001 (first entry)  
 DE Human cDNA encoding a novel secreted protein, Seq ID 284.  
 XX Human; immunosuppressive; antiarthritic; ss; antirheumatic;  
 KW cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnary; secreted protein; rheumatoid arthritis;  
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.

XX OS Homo sapiens.  
XX PN W0200155322-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01341.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX  
 DR WPI; 2001-488783/53.  
 DR P-PSDB; AAU16118.  
 XX  
 XX New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 PS Claim 1; SEQ ID NO 284; 980pp; English.  
 XX  
 XX The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection, and many others  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence encodes a novel secreted protein of the invention.

Query Match 81.1%; Score 15.4; DB 22; Length 970;  
 Best Local Similarity 94.1%; Pred. No. 2.1e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGGC 18  
 DB 930 GGGGATCACATGGTGGC 946

RESULT 14  
 AAD28130/C  
 ID AAD28130 standard; DNA; 2286 BP.  
 XX  
 AC AAD28130;  
 XX  
 DT 07-MAY-2002 (first entry)  
 XX  
 DE Soluble: secreted endopeptidase (SEP) consensus DNA.  
 XX  
 KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
 KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
 KW FSAD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
 XX  
 OS Homo sapiens.  
 OS Mus sp.  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1664..2286  
 FT /\*tag= a  
 FT /note= "Encodes catalytic domain"  
 XX  
 PN WO200206492-A1.  
 XX

PD 24-JAN-2002.  
 XX  
 PF 16-JUL-2001; 2001WO-IB01263.  
 XX  
 PR 14-JUL-2000; 2000GB-0017387.  
 XX  
 PA (PFIZ ) PFIZER LTD.  
 PA (PFIZ ) PFIZER INC.  
 XX  
 PI Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
 XX  
 DR WPI; 2002-155042/20.  
 XX  
 XX An isolated and/or purified nucleic acid encoding a human soluble.  
 PT secreted endopeptidase which is useful for treating sexual dysfunction,  
 PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
 XX female sexual arousal disorder -  
 PS Disclosure; Fig 6; 167pp; English.  
 XX  
 CC The invention relates to an isolated and/or purified nucleic acid  
 CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
 CC antibody and the compound which inhibits or selectively inhibits the  
 CC human SEP protein are useful in the manufacture of a medicament for  
 CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
 CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
 CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
 CC also useful for treating the above disorders and other disorders such  
 CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
 CC hypodactile sexual desire disorder. The present sequence is SEP consensus  
 CC DNA sequence found in human, mouse and rat.  
 XX  
 SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
 Query Match 81.1%; Score 15.4; DB 24; Length 2286;  
 Best Local Similarity 94.1%; Pred. No. 2.2e+02;  
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGATCACATGGTGGC 19  
 DB 363 GGGATCACATGGTGGC 347

RESULT 15  
 ABL28854/C  
 ID ABL28854 standard; DNA; 4183 BP.  
 XX  
 AC ABL28854;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 38035.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions

XX Claim 1; SEQ ID NO 38035; 21pp + Sequence Listing; English.

XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 4183 BP; 1201 A; 832 C; 886 G; 1264 T; 0 other;

Query Match 81.1%; Score 15.4; DB 23; Length 4183;  
Best Local Similarity 94.1%; Pred. No. 2.3e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGATCAGTGTGCGG 19

Db 1090 GGGATCTCATGTGCGG 1074  
||||| |||||

Search completed: July 8, 2003, 02:18:54  
Job time : 120.376 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 26.2293 Seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-8

Perfect score: 19

Sequence: 1 cggggatcacatggtgcg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0', Gapext 1.0.

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
C 1	15.8	83.2	504	US-09-624-390-3
C 2	15.8	83.2	1107	US-09-624-390-1
C 3	14.8	77.9	1679	US-08-676-882-1
C 4	14.8	77.9	1912	US-08-270-013B-1
C 5	14.8	77.9	1912	US-08-838-418-1
C 6	14.8	77.9	1977	US-08-272-255-17
C 7	14.8	77.9	1977	PCT-US95-08565-17
C 8	14.8	77.9	152331	US-09-128-155-16
C 9	14.8	77.9	176373	US-09-128-155-17
C 10	14.2	74.7	32	US-08-465-687A-7
C 11	14.2	74.7	32	US-09-030-970-7
C 12	14.2	74.7	1847	US-09-149-476-287
C 13	14.2	74.7	3366	US-08-469-802B-1
C 14	14.2	74.7	3366	US-08-267-803B-1
C 15	14.2	74.7	4214	US-09-221-017B-293
C 16	14.2	74.7	6830	US-08-822-445-1
C 17	14.2	74.7	6830	US-09-396-540-1
C 18	14.2	74.7	10660	US-08-267-803B-8
C 19	14.2	74.7	10660	US-09-041-886-16
C 20	14.2	74.7	4403765	US-09-103-840A-2
C 21	14.2	74.7	4411529	US-09-103-840A-1
C 22	13.8	72.6	649	US-09-470-191-5
C 23	13.8	72.6	1432	US-09-183-861-73
C 24	13.8	72.6	1432	US-09-022-765-73
C 25	13.8	72.6	1847	US-09-149-476-149
C 26	13.8	72.6	4550	US-09-103-663-35
C 27	13.8	72.6	28720	US-09-341-587-7

28	13.8	72.6	112132	4	US-09-741-150-3	Sequence 3, Appli
C 29	13.8	72.6	246240	2	US-08-724-394A-20	Sequence 20, Appli
C 30	13.8	72.6	246240	2	US-08-724-394A-21	Sequence 21, Appli
C 31	13.8	72.6	246240	2	US-08-724-394A-22	Sequence 22, Appli
C 32	13.4	70.5	252	4	US-08-905-223-232	Sequence 232, App
C 33	13.4	70.5	549	1	US-08-571-643A-1	Sequence 1, Appli
C 34	13.4	70.5	549	1	US-08-439-747A-1	Sequence 1, Appli
C 35	13.4	70.5	549	2	US-08-440-409B-1	Sequence 1, Appli
C 36	13.4	70.5	549	2	US-08-853-623D-16	Sequence 16, Appli
C 37	13.4	70.5	567	1	US-08-439-747A-3	Sequence 3, Appli
C 38	13.4	70.5	567	2	US-08-440-409B-3	Sequence 3, Appli
C 39	13.4	70.5	603	1	US-08-439-747A-4	Sequence 4, Appli
C 40	13.4	70.5	603	2	US-08-440-409B-4	Sequence 4, Appli
C 41	13.4	70.5	630	1	US-08-571-643A-2	Sequence 2, Appli
C 42	13.4	70.5	630	2	US-08-439-747A-5	Sequence 5, Appli
C 43	13.4	70.5	630	2	US-08-440-409B-5	Sequence 5, Appli
C 44	13.4	70.5	702	1	US-08-439-747A-7	Sequence 7, Appli
C 45	13.4	70.5	702	2	US-08-440-409B-7	Sequence 7, Appli

## ALIGNMENTS

RESULT 1  
US-09-624-390-3/c  
; Sequence 3, Application US/09624390  
; Patent No. 6372459  
; GENERAL INFORMATION:  
; APPLICANT: VERSECK, STEFAN  
; APPLICANT: KULA, MARIA-REGINA  
; APPLICANT: BOMMARIUS, ANDREAS  
; APPLICANT: DRAUZ, KARLHEINZ  
; TITLE OF INVENTION: N-ACTEYL AMINO ACID RACEMASE  
; FILE REFERENCE: 192535USO  
; CURRENT APPLICATION NUMBER: US/09/624,390  
; PRIORITY FILING DATE: 2000-07-27  
; PRIOR APPLICATION NUMBER: DE 19935268.2  
; PRIOR FILING DATE: 1999-07-27  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 504  
; TYPE: DNA  
; ORGANISM: Amycolatopsis orientalis  
US-09-624-390-3

Query Match 83.2%; Score 15.8; DB 4; Length 504;  
Best Local Similarity 89.5%; Pred. No. 19;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGATCACATGCGCG 19  
||||||| ||||| |||  
DB 220 CGGGATCAGATGCTCCG 202

RESULT 2  
US-09-624-390-1/c  
; Sequence 1, Application US/09624390  
; Patent No. 6372459  
; GENERAL INFORMATION:  
; APPLICANT: VERSECK, STEFAN  
; APPLICANT: KULA, MARIA-REGINA  
; APPLICANT: BOMMARIUS, ANDREAS  
; APPLICANT: DRAUZ, KARLHEINZ  
; TITLE OF INVENTION: N-ACTEYL AMINO ACID RACEMASE  
; FILE REFERENCE: 192535USO  
; CURRENT APPLICATION NUMBER: US/09/624,390  
; PRIORITY FILING DATE: 2000-07-27  
; PRIOR APPLICATION NUMBER: DE 19935268.2  
; PRIOR FILING DATE: 1999-07-27  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1

LENGTH: 1107  
TYPE: DNA  
ORGANISM: Ancycolatopsis orientalis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1107)  
US-09-624-390-1

Query Match 83.2%; Score 15.8; DB 4; Length 1107;  
Best Local Similarity 89.5%; Pred. No. 20;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCATGTCGCGC 19  
|||||  
Db 220 CGGGGATCATGTCGCGC 202

RESULT 3  
US-08-676-882-1  
; Sequence 1, Application US/08676882  
; Patent No. 6100241  
; GENERAL INFORMATION:  
; APPLICANT: Kok, Jacobus Johannes  
; APPLICANT: van den Boogaart, Paul  
; APPLICANT: Vermeulen, Arnoldus Nicolaas  
; TITLE OF INVENTION: Coccidiosis poultry vaccine  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Akzo No. 6100241el Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676,882  
; FILING DATE: 03-JUL-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gormley, Mary E.  
; REGISTRATION NUMBER: 34,409  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 258-5200  
; TELEFAX: (301) 977-0847  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1679 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Elmeria acervulina  
; DEVELOPMENTAL STAGE: Schizont  
; IMMEDIATE SOURCE:  
; CLONE: EASC2\_1  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 280..1269  
; OTHER INFORMATION: /function= "Elmeria lactate  
; OTHER INFORMATION: dehydrogenase"  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..51  
; OTHER INFORMATION: /label= pbluescriptII

NAME/KEY: misc\_feature  
LOCATION: 1624..1679  
OTHER INFORMATION: /label= pbluescriptII  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 45..54  
OTHER INFORMATION: /label= EcoRI-linker  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1621..1630  
OTHER INFORMATION: /label= EcoRI-linker  
US-08-676-882-1

Query Match 77.9%; Score 14.8; DB 3; Length 1679;  
Best Local Similarity 88.9%; Pred. No. 68;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCATGTCGCGC 18  
|||||  
Db 840 CGGGGATCATGTCGCGC 857

RESULT 4  
US-08-270-013B-1  
; Sequence 1, Application US/08270013B.  
; Patent No. 5686294  
; GENERAL INFORMATION:  
; APPLICANT: Sodabe et al.  
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE  
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 61601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/270,013B  
; FILING DATE: 01-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 164701/1993  
; FILING DATE: 02-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Robert F.  
; REGISTRATION NUMBER: 27555  
; REFERENCE/DOCKET NUMBER: 62321  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; TELEX: (25)3533  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1912 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus stearothermophilus  
; STRAIN: ATCC12016  
; US-08-270-013B-1  
Query Match 77.9%; Score 14.8; DB 1; Length 1912;

Best Local Similarity 88.9%; Pred. No. 69;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGCCG 19  
DB 1223 GCGGATGACATGGTGCCG 1240

## RESULT 5

US-08-838-418-1  
; Sequence 1, Application US/08838418  
; Patent No. 5744342  
; GENERAL INFORMATION:  
; APPLICANT: Sogabe et al.  
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/838,418  
; FILING DATE: 17-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/270,013  
; FILING DATE: 01-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 164701/1993  
; FILING DATE: 02-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hoover, Allen E.  
; REGISTRATION NUMBER: 37354  
; REFERENCE/DOCKET NUMBER: 78339  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; TELEX: (25)3533  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1912 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Bacillus stearothermophilus  
; STRAIN: ATCC12016  
; US-08-838-418-1

Query Match 77.9%; Score 14.8; DB 1; Length 1912;  
Best Local Similarity 88.9%; Pred. No. 69;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGCCG 19  
DB 1223 GCGGATGACATGGTGCCG 1240

## RESULT 6

US-08-272-255-17

; Sequence 17, Application US/08272255  
; Patent No. 5824859  
; GENERAL INFORMATION:  
; APPLICANT: Cashmore, Anthony R.  
; APPLICANT: Ahmad, Margaret  
; APPLICANT: Lin, Chentao  
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
; TITLE OF INVENTION: Using the Same  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859rls  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/272,255  
; FILING DATE: 08-JUL-1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leary Ph.D.; Kathryn  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: UPN-1795  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1977 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; US-08-272-255-17

Query Match 77.9%; Score 14.8; DB 1; Length 1977;  
Best Local Similarity 88.9%; Pred. No. 69;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGGTGCC 18  
DB 1304 CGGGGATCACATGGTGCC 1321

## RESULT 7

PCT-US95-08565-17  
; Sequence 17, Application PC/TUS9508565  
; GENERAL INFORMATION:  
; APPLICANT: Cashmore, Anthony R.  
; APPLICANT: Ahmad, Margaret  
; APPLICANT: Lin, Chentao  
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
; TITLE OF INVENTION: Using the Same  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

Query Match 77.9%; Score 14.8; DB 1; Length 1912;  
Best Local Similarity 88.9%; Pred. No. 69;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGCCG 19  
DB 1223 GCGGATGACATGGTGCCG 1240

APPLICATION NUMBER: PCT/US95/08565  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA: US 08/272,255  
FILING DATE: 08-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: UPN-1795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1977 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
PCT-US95-08565-17

Query Match 77.9%; Score 14.8; DB 5; Length 1977;  
Best Local Similarity 88.9%; Pred. No. 69;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGGGGATCAGATGGTGCC 18  
|| ||||| |||||  
Db 1304 CGAGGATCAGATGGTGCC 1321

RESULT 8  
US-09-128-155-16  
Sequence 16, Application US/09128155  
Patent No. 6117654  
GENERAL INFORMATION:  
APPLICANT: Pan, Yang  
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
FILE REFERENCE: 09404/052001  
CURRENT APPLICATION NUMBER: US/09/128,155  
CURRENT FILING DATE: 1998-08-03  
EARLIER APPLICATION NUMBER: US 60/091,650  
EARLIER FILING DATE: 1998-07-02  
EARLIER APPLICATION NUMBER: US 60/054,646  
EARLIER FILING DATE: 1997-08-04  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 16  
LENGTH: 152331  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(152331)  
OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-16

Query Match 77.9%; Score 14.8; DB 3; Length 152331;  
Best Local Similarity 88.9%; Pred. No. 89;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGGGGATCAGATGGTGCC 18  
|| ||||| |||||  
Db 65525 CTGGGATCAGCTGGTGCC 65542

RESULT 9  
US-09-128-155-17/c  
Sequence 17, Application US/09128155  
Patent No. 6117654  
GENERAL INFORMATION:  
APPLICANT: Pan, Yang

TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
FILE REFERENCE: 09404/052001  
CURRENT APPLICATION NUMBER: US/09/128,155  
CURRENT FILING DATE: 1998-08-03  
EARLIER APPLICATION NUMBER: US 60/091,650  
EARLIER FILING DATE: 1998-07-02  
EARLIER APPLICATION NUMBER: US 60/054,646  
EARLIER FILING DATE: 1997-08-04  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 17  
LENGTH: 176373  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(176373)  
OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-17

Query Match 77.9%; Score 14.8; DB 3; Length 176373;  
Best Local Similarity 88.9%; Pred. No. 90;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGGGGATCAGATGGTGCC 18  
|| ||||| |||||  
Db 69425 CTGGGATCAGCTGGTGCC 69408

RESULT 10  
US-08-465-687A-7/c  
Sequence 7, Application US/08465687A  
Patent No. 5750370  
GENERAL INFORMATION:  
APPLICANT: LI, ET AL.  
TITLE OF INVENTION: Human Endothelin Receptor  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,687A  
FILING DATE: 6 JUNE 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/11843  
FILING DATE: 17 OCT 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MULLINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-322 (PF137)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: Oligonucleotide  
US-08-465-687A-7



EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 669  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 312  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048, 974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056, 886  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 877  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057, 761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047, 595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 590  
EARLIER FILING DATE: 1997-05-23

EARLIER APPLICATION NUMBER: 60/047, 594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043, 578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047, 501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043, 670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056, 632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048, 964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057, 650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056, 884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057, 669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049, 610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061, 060  
EARLIER FILING DATE: 1997-10-02

Query Match 74.7%; Score 14.2; DB 4; Length 1847;  
Best Local Similarity 84.2%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGGATCACATGGTGGC 19  
Db 94 CGGGAGAGATGGTGGC 112

## RESULT 13

US-08-469-802B-1/G  
Sequence 1, Application US/08469802B  
Patent No. 5741645  
GENERAL INFORMATION:  
APPLICANT: Orr, Harry T.  
APPLICANT: Rarum, Laura P.W.  
APPLICANT: Chung, Ming-yl  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
Patent No. 5741645  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mueeting, Raasch, Gebhardt & Schwappach, P.A.  
STREET: 119 No. 5741645th Fourth Street, Suite 203  
CITY: Minneapolis

STATE: MN  
COUNTRY: USA  
ZIP: 55401  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,802B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muetting, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110.00030101  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1225  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3366 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-469-802B-1

Query Match 74.7%; Score 14.2; DB 1; Length 3366;  
Best Local Similarity 84.2%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGGGATCATCGTGCCG 19  
|||||  
Db 2628 CGGGGATCACCAGGTCTG 2610

RESULT 14  
US-08-267-803B-1/c  
Sequence 1, Application US/08267803B  
Patent No. 5834183  
GENERAL INFORMATION:  
APPLICANT: Orr, Harry T.  
APPLICANT: Ranum, Laura P.W.  
APPLICANT: Chung, Ming-yi  
APPLICANT: Zoghbi, Huda Y.  
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
Patent No. 5834183  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Muetting, Raasch, Gebhardt & Schwappach, P.A.  
STREET: P.O. Box 581415  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55458-1415  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/267,803B  
FILING DATE: 28-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McCoimack, Myra H.  
REGISTRATION NUMBER: 36,602  
REFERENCE/DOCKET NUMBER: 110.00030120  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3366 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-267-803B-1

Query Match 74.7%; Score 14.2; DB 2; Length 3366;  
Best Local Similarity 84.2%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGGGATCATCGTGCCG 19  
|||||  
Db 2628 CGGGGATCACCAGGTCTG 2610

RESULT 15  
US-09-221-017B-293  
Sequence 293, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H.  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 293:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4214 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: 1...4214

US-09-221-017B-293

Query Match 74.7%; Score 14.2; DB 4; Length 4214;  
Best Local Similarity 84.2%; Pred. No. 1.5e+02;  
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGGTGCCG 19

|| ||||| ||||| |||||

Db 3737 CGTCGATCACATGGAGCCG 3755

Search completed: July 8, 2003, 09:31:08

Job time : 32.2793 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 109.737 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-8

Perfect score: 19

Sequence: 1 cggggatcacatggtgccg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/ECTUS\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.8	83.2	1107	10	US-09-973-765-1
C 2	15.8	83.2	1107	10	US-09-973-712-1
C 3	15.4	81.1	970	10	US-09-764-864-284
5	14.8	77.9	240	10	US-09-923-876-2432
C 6	14.8	77.9	491	9	US-10-079-623-81
C 7	14.8	77.9	831	10	US-09-867-701-10019
C 8	14.8	77.9	6459	9	US-10-029-217A-6
C 9	14.8	77.9	20987	9	US-09-764-891-8727
10	14.8	77.9	24533	9	US-09-764-868-1349
C 11	14.8	77.9	152331	9	US-10-095-407-16
C 12	14.8	77.9	176373	9	US-10-095-407-17
C 13	14.4	75.8	97	10	US-09-864-761-27728
C 14	14.4	75.8	460	10	US-09-864-761-11096
15	14.4	75.8	546	10	US-09-815-242-6527
16	14.4	75.8	2051	10	US-09-917-800A-1325
17	14.4	75.8	3129	9	US-10-161-803-24
18	14.4	75.8	3172	9	US-10-013-477-6
19	14.4	75.8	14962	9	US-10-079-854-244
			14962	10	US-09-764-878-244

```

c 20 14.2 74.7 32 9 US-10-292-525-7 Sequence 7, Appli
c 21 14.2 74.7 309 10 US-09-867-701-10553 Sequence 10553, A
c 22 14.2 74.7 458 9 US-09-918-995-25291 Sequence 25291, A
c 23 14.2 74.7 484 9 US-09-918-995-22699 Sequence 22699, A
c 24 14.2 74.7 500 9 US-09-918-995-20655 Sequence 20655, A
c 25 14.2 74.7 502 9 US-09-918-995-31928 Sequence 31928, A
c 26 14.2 74.7 504 9 US-10-156-761-6460 Sequence 6460, Ap
c 27 14.2 74.7 750 10 US-09-910-943-365 Sequence 365, App
c 28 14.2 74.7 771 10 US-09-910-943-152 Sequence 152, App
c 29 14.2 74.7 915 9 US-10-156-761-2013 Sequence 2013, Ap
c 30 14.2 74.7 1150 9 US-10-239-420-13 Sequence 13, Appl
c 31 14.2 74.7 1203 10 US-09-974-300-996 Sequence 996, App
c 32 14.2 74.7 1411 9 US-10-037-270-535 Sequence 535, App
c 33 14.2 74.7 1518 9 US-10-116-821-13 Sequence 13, Appl
c 34 14.2 74.7 1518 9 US-10-117-283-13 Sequence 13, Appl
c 35 14.2 74.7 1620 9 US-10-116-175-1 Sequence 1, Appl
c 36 14.2 74.7 1717 9 US-10-197-666A-133 Sequence 133, App
c 37 14.2 74.7 1725 9 US-10-156-761-7096 Sequence 7096, A
c 38 14.2 74.7 1847 9 US-09-809-391-287 Sequence 287, App
c 39 14.2 74.7 2118 9 US-10-156-761-290 Sequence 290, App
c 40 14.2 74.7 2232 9 US-09-764-891-5479 Sequence 5479, Ap
c 41 14.2 74.7 2232 9 US-09-764-891-10206 Sequence 10206, A
c 42 14.2 74.7 2232 9 US-10-205-428-1005 Sequence 1005, A
c 43 14.2 74.7 2893 9 US-10-017-273A-4 Sequence 4, Appli
c 44 14.2 74.7 2893 10 US-09-905-846-1 Sequence 1, Appli
c 45 14.2 74.7 2975 9 US-10-017-273A-5 Sequence 5, Appli

```

#### ALIGNMENTS

##### RESULT 1

```

US-09-973-765-1/c
; Sequence 1, Application US/09973765
; Patent No. US20020090684A1
; GENERAL INFORMATION:
; APPLICANT: BOMMARIUS, ANDREAS
; APPLICANT: DRAUZ, KARLHEINZ
; APPLICANT: VERSECK, STEFAN
; TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF AMINO ACIDS
; FILE REFERENCE: 214381US-10757-9350-0-X
; CURRENT APPLICATION NUMBER: US/09/973,765
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: DE 100 50 123.0
; PRIOR FILING DATE: 2000-10-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1107)
; OTHER INFORMATION:
US-09-973-765-1

```

Query Match 83.2%; Score 15.8; DB 10; Length 1107;

Best Local Similarity 89.5%; Pred. No. 67;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGGATCACATGTCGCCG 19

||||||| ||||| |||

DB 220 CGGGGATCACATGTCGCCG 202

##### RESULT 2

```

US-09-973-712-1/c
; Sequence 1, Application US/09973712
; Patent No. US20020106752A1
; GENERAL INFORMATION:
; APPLICANT: BOMMARIUS, ANDREAS
; APPLICANT: DRAUZ, KARLHEINZ

```

```
; APPLICANT: VERSECK, STEFAN
; APPLICANT: KULA, MARIA-REGINA
; TITLE OF INVENTION: ACETYL AMINO ACID RACEMASE FROM AMYCOLATOPSIS ORIENTALIS FOR RACE
; FILE REFERENCE: 214382USOX
; CURRENT APPLICATION NUMBER: US/09/973.712
; PRIOR FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: DE 10050124.9
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: Amycolatopsis orientalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1107)
; OTHER INFORMATION:
US-09-973-712-1

Query Match      83.2%; Score 15.8; DB 10; Length 1107;
Best Local Similarity 89.5%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGGATCATCATGGTGCCG 19
Db 220 CGGGATCATCATGGTGCCG 202

RESULT 3
US-09-764-864-284
; Sequence 284, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764.864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 284
; LENGTH: 970
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-284

Query Match      81.1%; Score 15.4; DB 10; Length 970;
Best Local Similarity 94.1%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGATCATCATGGTGCC 18
Db 930 GGGATCATCATGGTGCC 946

RESULT 4
US-09-923-876-2432
; Sequence 2432, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Laigudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923.876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
```

```
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 2432
; LENGTH: 240
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700160596H1
; NAME/KEY: unsure
; LOCATION: 43, 174, 183-184, 197-198
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-2432

Query Match      77.9%; Score 14.8; DB 10; Length 240;
Best Local Similarity 88.9%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGGATCATCATGGTGCCG 19
Db 112 GAGGACCATCATGGTGCCG 129

RESULT 5
US-10-079-623-81
; Sequence 81, Application US/10079623
; Patent No. US20020169302A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; FILE REFERENCE: 11000.1044c3
; CURRENT APPLICATION NUMBER: US/10/079,623
; CURRENT FILING DATE: 2002-02-19
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 81
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Bovine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(491)
; OTHER INFORMATION: n = A,T,C or G
US-10-079-623-81

Query Match      77.9%; Score 14.8; DB 9; Length 491;
Best Local Similarity 88.9%; Pred. No. 2.2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGGATCATCATGGTGCCG 19
Db 56 GGGATCATCATGGTGCCG 73

RESULT 6
US-09-867-701-10019/c
; Sequence 10019, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
```

```

RESULT 8
US-09-764-891-8727/c
; Sequence 8727, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 8727
; LENGTH: 20987
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (11074)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-8727

```

Qy 1 CGGGATCACATGGTGCC. 18  
| | | | | | | | | |  
Db 65525 CTGGGATCACCTGGTGCC 65542

## RESULT 11

US-10-095-407-17/c  
; Sequence 17, Application US/10095407  
; Patent No. US20020164330A1  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/10/095,407  
; CURRENT FILING DATE: 2002-03-11  
; PRIOR APPLICATION NUMBER: US 60/091,650  
; PRIOR FILING DATE: 1998-07-02  
; PRIOR APPLICATION NUMBER: US 60/054,646  
; PRIOR FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 176373.  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(176373)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-095-407-17

Query Match 77.9%; Score 14.8; DB 9; Length 176373;  
Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGGGATCATGTGTC 18  
Db 69425 CTGGATCACCTGTGCC 69408  
|||||||

## RESULT 12

US-09-864-761-27728/c  
; Sequence 27728, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 27728  
; LENGTH: 97  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC002299.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.92  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.87  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.86  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.83  
; OTHER INFORMATION: EST\_HUMAN HIT: BE379328.1, EVALUATE 7.60e-02  
US-09-864-761-27728

Query Match 75.8%; Score 14.4; DB 10; Length 97;  
Best Local Similarity 93.8%; Pred. No. 3.3e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 GGATCATGTGTCGC 19  
Db 87 GGATCATGTGTCAG 72  
|||||||

## RESULT 13

US-09-864-761-11096/c  
; Sequence 11096, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30

Search completed: July 9, 2003, 02:21:52

Job time : 109.787 secs

---

```

RESULT 15
BQ984595
LOCUS       BQ984595               515 bp    mRNA    linear    EST 21-AUG-2002
DEFINITION  QGE3C05.Y9.ab1 QG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
            QGE3C05, mRNA sequence.
ACCESSION   BQ984595
VERSION     BQ984595.1  GI:22402120
KEYWORDS    EST.
SOURCE      BQ984595.1
            Lactuca sativa.
            Lactuca sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridaceae; euasterids II; Asterales; Asteraceae; Lactuceae;
            Lactuca.
REFERENCE   1 (bases 1 to 515)
AUTHORS    Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
            Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
            ,P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
            Church,S., Jackson,L. and Bradford,K.
            Lettuce and Sunflower ESTs from the Compositae Genome Project
            http://compgenomics.ucdavis.edu/
            unpublished (2002)
JOURNAL    Contact: Alexander Kozik [R.W.Michelmore]
COMMENT    Department of Vegetable Crops, R.W.Michelmore Lab
            University of California at Davis (UCD)
            Asmundson Hall, UCD, Davis, CA 95616, USA
            Tel: 1-(530)-742-1742
            Fax: 1-(530)-752-9659
            Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
            belongs to contig QG_CA.Contig3484, see http://cgpdb.ucdavis.edu/
            for details.
            Plate: QGE3. row: c column: 05.
            Location/Qualifiers
                1..515
                /organism="Lactuca sativa"
                /cultivar="L.serriola"
                /db_xref="taxon:4236"
                /clone="QGE3C05"
                /clone_lib="QG_EFGHJ lettuce serriola"
                /lab_host="E.coli"
                /note="Vector: pBRCDNASfiAB; The library was constructed
                from 10 different sources of RNA from a single genotype.
                Separate cDNAs were generated using primers that
                incorporated unique 5' and 3' tags to distinguish each
                source of RNA. cDNAs were then pooled, size-fractionated,
                directionally cloned into a custom medium-copy vector and
                transformations made with four size classes to minimize
                size bias. Details of each source of RNA and library
                construction can be obtained at http://cgpdb.ucdavis.edu/
                TAG_LIB-QG_EFGHJ lettuce serriola
                TAG_TISSUE=leaves dark grow
                TAG_SEQ=GCTAGTCGGG"
BASE COUNT      98 a      170 g      158 t
ORIGIN
Query Match      86.3%; Score 16.4; DB 14; Length 515;
Best Local Similarity 94.4%; Pred. NO. 1.1e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      2 GGGGATCACATGGTCCG 19
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      44 GAGGATCACATGGTCCG 61

```

Search completed: July 8, 2003, 09:21:26  
 Job time : 968.404 secs



JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:942781  
High quality sequence stop: 421.

FEATURES  
source

Location/Qualifiers  
1. .503  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1515929"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
ORIGIN

133 a 132 c 107 g 131 t

Query Match

Best Local Similarity 86.3%; Score 16.4; DB 10; Length 503;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATCATCATGGTGCC 18

1 |||||

Db 420 CAGGGATCATCATGGTGCC 403

RESULT 13  
AW986364/C

LOCUS

uf79g03.x1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
IMAGE:1528372 3', mRNA sequence.

ACCESSION

AW986364.1 GI:8180983

VERSION

EST.

KEYWORDS

house mouse.

SOURCE

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 504)  
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ccgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other ESTs: uf79g03.y1

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:944472

High quality sequence stop: 450.

FEATURES  
source

Location/Qualifiers  
1. .504  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1528372"  
/clone\_lib="Soares\_mammary\_gland\_NMLMG"  
/sex="female (lactating)"  
/tissue\_type="mammary gland"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from mammary  
gland tissue from a lactating female, and was then primed

with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT7T3 vector. Library is normalized. Library  
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

133 a 132 c 108 g 131 t

ORIGIN

Query Match

Best Local Similarity 86.3%; Score 16.4; DB 10; Length 504;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGATCATCATGGTGCC 18

1 |||||

Db 420 CAGGGATCATCATGGTGCC 403

RESULT 14

AA214179

LOCUS

AA214179 514 bp mRNA linear EST 12-MAR-1998  
z990607.r1 Stratagene hnt neuron (#937233) Homo sapiens cDNA clone  
IMAGE:649284 5' similar to contains Alu repetitive element; contains  
element MER22 repetitive element; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA214179.1 GI:1812859  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 514)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 2292 Std Error: 0.00  
Seq primer: -28ml3 revl ET from Amersham  
High quality sequence stop: 442.

FEATURES  
source

1. .514  
/organism="Homo sapiens"  
/db\_xref="GDB:5278563"  
/db\_xref="taxon:9606"  
/clone="IMAGE:649284"  
/clone\_lib="Stratagene hnt neuron (#937233)"  
/dev\_stage="hnt neurons"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
XhoI; Cloned unidirectionally. Primer: Oligo dT.  
Differentiated, post mitotic hnt neurons. Average insert  
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'  
GAATTCGGCAG 3' -3' adaptor sequence: 5'  
CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT

125 a 110 c 162 g 114 t 3 others

ORIGIN

Query Match

Best Local Similarity 86.3%; Score 16.4; DB 9; Length 514;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGATCATCATGGTGCC 19

|||||

Db 315 GGGATCATCATGGTGCC 332



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 963.254 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-8

Perfect score: 19

Sequence: 1 cggggatcacatggtccg 19

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- EST:\*\*
- 1: em\_estba:\*\*
  - 2: em\_esthum:\*\*
  - 3: em\_estlin:\*\*
  - 4: em\_estmdu:\*\*
  - 5: em\_estov:\*\*
  - 6: em\_estpl:\*\*
  - 7: em\_estro:\*\*
  - 8: em\_htc:\*\*
  - 9: gb\_estli:\*\*
  - 10: gb\_estc2:\*\*
  - 11: gb\_htc:\*\*
  - 12: gb\_est3:\*\*
  - 13: gb\_est4:\*\*
  - 14: gb\_est5:\*\*
  - 15: em\_estfun:\*\*
  - 16: em\_estom:\*\*
  - 17: gb\_gss:\*\*
  - 18: em\_gss\_hum:\*\*
  - 19: em\_gss\_inv:\*\*
  - 20: em\_gss\_pln:\*\*
  - 21: em\_gss\_vrt:\*\*
  - 22: em\_gss\_fun:\*\*
  - 23: em\_gss\_mam:\*\*
  - 24: em\_gss\_mus:\*\*
  - 25: em\_gss\_other:\*\*
  - 26: em\_gss\_pro:\*\*
  - 27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	89.5	325	12	BG055194 nad03e01
2	16.4	86.3	122	12	BF992713 CM1-GN016
3	16.4	86.3	242	14	BQ990877 OGF21F15
4	16.4	86.3	383	10	BQ990877 uF79g03.y
5	16.4	86.3	391	10	BE448826 uc88c12.y
6	16.4	86.3	410	10	BB751033 BB751033

7	16.4	86.3	413	14	BQ980809
8	16.4	86.3	414	10	BB792684
9	16.4	86.3	432	10	AW492539
10	16.4	86.3	457	12	BF468707
11	16.4	86.3	467	10	AW823984
12	16.4	86.3	503	10	AW823459
13	16.4	86.3	504	10	AW86364
14	16.4	86.3	514	9	AA214179
15	16.4	86.3	515	14	BQ984595
16	16.4	86.3	525	17	AQ118971
17	16.4	86.3	557	14	BQ10846
18	16.4	86.3	565	14	BQ983268
19	16.4	86.3	587	10	AW557704
20	16.4	86.3	594	17	BH727026
21	16.4	86.3	646	12	BQ075949
22	16.4	86.3	648	14	BQ868940
23	16.4	86.3	671	10	BB089035
24	16.4	86.3	714	14	BQ043003
25	16	84.2	372	9	AA893502
26	16	84.2	556	12	BE820119
27	15.8	83.2	337	14	BQ019635
28	15.8	83.2	414	17	AQ067581
29	15.8	83.2	462	17	AQ067260
30	15.8	83.2	472	12	BG609896
31	15.8	83.2	525	17	AQ356344
32	15.8	83.2	561	17	AQ356342
33	15.8	83.2	619	17	AQ308594
34	15.8	83.2	622	9	AL680669
35	15.8	83.2	651	17	AQ353259
36	15.8	83.2	667	10	AW676437
37	15.8	83.2	667	12	BF203825
38	15.8	83.2	757	13	BF869775
39	15.8	83.2	908	12	BF530814
40	15.8	83.2	957	17	CNS043UD
41	15.8	83.2	981	14	BQ892809
42	15.8	83.2	1039	17	CNS040MC
43	15.8	83.2	1046	17	CNS040I9
44	15.4	81.1	207	10	BB608514
45	15.4	81.1	230	10	AV369289

## ALIGNMENTS

RESULT 1  
BG055194  
LOCUS  
DEFINITION  
nad03e01.x1 NCL\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:3432289 3', similar to contains Alu repetitive element; contains element MER22 repetitive element ;, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BG055194 325 bp mRNA linear EST 25-JAN-2001  
nad03e01.x1 NCL\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:3432289 3', similar to contains Alu repetitive element; contains element MER22 repetitive element ;, mRNA sequence.  
BG055194  
BG055194.1 GI:12512671  
EST.  
human  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 325)  
NCL\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov.  
Seq primer: -400P from Gibco.

FEATURES  
source

Location/Qualifiers  
1. 325  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3432289"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacla) with a modified polylinker; Plasmid DNA from the normalized library NCI\_CGAP\_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 985608-986759, 110192-1101959, and 1217928-1220615). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 65 a 108 c 71 g 81 t

## ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 325;  
Best Local Similarity 100.0%; Pred. NO. 4.7e+02;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGATCACATGGTGC 17

|||||

Db 131 CGGGATCACATGGTGC 147

## RESULT 2

## BF992713

## LOCUS

BF992713 122 bp mRNA linear EST 23-JAN-2001

DEFINITION CM1-GN0161-301000-517-h04 GN0161 Homo sapiens cDNA, mRNA sequence.

## ACCESSION

## BF992713

## VERSION

## EST

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## COMMENT

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?ti=CM1&t2=CM1-GN0161-301000-517-h04&t3=2000-10-30&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 9

High quality sequence stop: 122.

Location/Qualifiers

1. 122

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="GN0161"

/dev\_stage="Adult"

## FEATURES

## source

Location/Qualifiers

1. 122

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="GN0161"

/dev\_stage="Adult"

## BASE COUNT

## ORIGIN

## FEATURES

## source

## Location/Qualifiers

## 1. 242

## /organism="Lactuca sativa"

## /cultivar="L. serriola"

## /db\_xref="taxon:4236"

## /clone="QGF21F15"

## /clone\_lib="QG\_EFGHJ lettuce serriola"

## /lab\_host="E.coli"

## /note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG\_Lib-QG\_EFGHJ lettuce serriola TAG\_TISSUE-leaves dark grow TAG\_SEQ-GCTAGTCGGG"

## BASE COUNT

## 38 a

## 35 c

## 131 g

## 38 t

/note="Organ: placenta\_normal; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 30 a 21 c 37 g 34 t

## ORIGIN

Query Match 86.3%; Score 16.4; DB 12; Length 122;

Best Local Similarity 94.4%; Pred. NO. 6.9e+02;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGGATCACATGGTGC 19

|||||

Db 48 GGGGATCACATGGTGC 65

## RESULT 3

## BQ990877

## LOCUS

BQ990877 242 bp mRNA linear EST 21-AUG-2002

DEFINITION QGF21F15.yg.ab1 QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone

QGF21F15, mRNA sequence.

ACCESSION BQ990877

VERSION BQ990877.1 GI:22410412

KEYWORDS EST.

SOURCE Lactuca sativa.

ORGANISM Lactuca sativa

REFERENCE 1 (bases 1 to 242)

AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.

Lettuce and Sunflower ESTs from the Compositae Genome Project

http://compositae.ucdavis.edu/

Unpublished (2002)

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmundson Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig QG\_CA\_Contig3186, see http://cgdb.ucdavis.edu/

for details

Plate: QGF21 row: F column: 15.

Location/Qualifiers

1. 242

/organism="Lactuca sativa"

/cultivar="L. serriola"

/db\_xref="taxon:4236"

/clone="QGF21F15"

/clone\_lib="QG\_EFGHJ lettuce serriola"

/lab\_host="E.coli"

/note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at http://cgdb.ucdavis.edu/TAG\_Lib-QG\_EFGHJ lettuce serriola TAG\_TISSUE-leaves dark grow TAG\_SEQ-GCTAGTCGGG"

## BASE COUNT

## 38 a

## 35 c

## 131 g

## 38 t



Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watanishi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384 format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y., and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

## e mouse tissues.

## FEATURES Location/Qualifiers

1. 410  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone\_lib="G130203K05"  
/clone\_lib="RIKEN full-length enriched, pooled tissues,  
cerebellum, etc."  
/note="pooled tissues: (tissue\_type=cerebellum,  
dev\_stage=16 days neonate, sex=mixed),  
(tissue\_type=cerebellum, dev\_stage=0 day neonate,  
sex=mixed), (tissue\_type=hippocampus, dev\_stage=adult,  
sex=male), (tissue\_type=whole body, dev\_stage=9 days  
embryo, sex=mixed), (tissue\_type=lung, dev\_stage=13 days  
embryo, sex=mixed)."

BASE COUNT 94 a 108 c 108 g 100 t

## ORIGIN

Query Match 86.3%; Score 16.4; DB 10; Length 410;  
Best Local Similarity 94.4%; Pred. No. 9.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGGGGATCATCATGGTGGCC 18  
Db 10 CAGGGATCATCATGGTGGCC 27

## RESULT 7

## BQ980809

## LOCUS

BQ980809 413 bp mRNA linear EST 21-AUG-2002  
QGE12B10.yg.ab1 OG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
QGE12B10, mRNA sequence.

## ACCESSION

## BQ980809

## VERSION

## QY

## KEYWORDS

## EST.

## SOURCE

## Lactuca sativa

## ORGANISM

## Lactuca sativa

## REFERENCE

## AUTHORS

## 1 (bases 1 to 413)

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
Church, S., Jackson, L., and Bradford, K.  
Lettuce and Sunflower ESTs from the Compositae Genome Project  
<http://comgenomics.ucdavis.edu/>  
Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab

## TITLE

## JOURNAL

## COMMENT

University of California at Davis (UCD)  
Amundson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]  
belongs to contig OG\_CA\_Contig3186, see <http://cgdb.ucdavis.edu/>  
for details.  
Plate: OGE12 row: B column: 10.

## FEATURES

## source

## Location/Qualifiers

## 1. 413

## /organism="Lactuca sativa"

## /cultivar="L. serriola"

## /db\_xref="taxon:4236"

## /clone\_lib="QGE12B10"

## /clone\_lib="OG\_EFGHJ lettuce serriola"

## /lab\_host="E.coli"

/note="Vector: pBRCNASFIAB: The library was constructed  
from 10 different sources of RNA from a single genotype.  
Separate cDNAs were generated using primers that  
incorporated unique 5' and 3' tags to distinguish each  
source of RNA. cDNAs were then pooled, size-fractionated,  
directionally cloned into a custom medium-copy vector and  
transformations made with four size classes to minimize  
size bias. Details of each source of RNA and library  
construction can be obtained at <http://cgdb.ucdavis.edu/>  
TAG LIB=OG\_EFGHJ lettuce serriola  
TAG TISSUE=chemical induction  
TAG\_SEQ=TCGTAGCCGGG"

## BASE COUNT

## 68 a 52 c 222 g 71 t

## ORIGIN

Query Match 86.3%; Score 16.4; DB 14; Length 413;  
Best Local Similarity 94.4%; Pred. No. 9.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GGGGATCATCATGGTGGCC 19  
Db 157 GAGGATCATCATGGTGGCC 174

## RESULT 8

## BQ980809

## LOCUS

## DEFINITION

## BQ980809

## ACCESSION

## VERSION

## KEYWORDS

## EST.

## SOURCE

## house mouse

## ORGANISM

## Mus musculus

## REFERENCE

## AUTHORS

## 1 (bases 1 to 414)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hirose, T., Hirozane, T., Imotani, K., Ishii  
Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,  
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T.,  
Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,  
Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa  
A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,  
Watanishi, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 131.941 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-9

Perfect score: 21

Sequence: 1 ctaccccaagctgctgatag 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	AAZ28815	Rat membrane metal
2	21	100.0	2765	AAZ28810	Rat membrane metal
3	19.4	92.4	2286	AAZ28130	Soluble secreted e
C 4	17.8	84.8	1177	AAZ48434	Arabidopsis thalia
C 5	17.8	84.8	1218	AAZ39055	Arabidopsis thalia
6	17.8	84.8	2925	AAZ37763	CDNA encoding neut
7	16.8	80.0	1521	ABQ55050	Human ovarian anti
8	16.8	80.0	2026	ABK35251	Human cDNA encodin
9	16.8	80.0	2835	AAV58755	Human secreted pro

10	16.8	80.0	3328	24	ABK84434	Human cDNA differe
11	16.8	80.0	3610	23	AAS84900	DNA encoding novel
12	16.8	80.0	4424	23	AAS84899	DNA encoding novel
13	16.8	80.0	7606	22	AAS28366	Genomic sequence #
14	16.8	80.0	7608	22	AAS28367	Human musculoskele
15	16.8	80.0	12771	22	AAL37558	Genomic sequence #
16	16.8	80.0	17498	22	AAS28364	Genomic sequence #
17	16.8	80.0	32146	22	AAS28363	Human immune/hema
18	16.8	80.0	34435	22	AAK76172	Human cDNA differe
C 19	16.8	80.0	165199	24	AAK83460	DNA encoding a coc
20	16.4	78.1	629	20	AAK04758	Oligonucleotide us
21	16.2	77.1	40	22	AAF89758	Nucleic acid fragm
22	16.2	77.1	180	22	AAF89764	Nucleotide sequenc
23	16.2	77.1	2076	22	AAF89737	Human metalloprote
24	16.2	77.1	2232	24	AAS97186	Nucleotide sequenc
25	16.2	77.1	2262	22	AAF89739	Human protease, PR
26	16.2	77.1	2318	24	AAD30580	Nucleotide sequenc
27	16.2	77.1	2340	22	AAF89738	Human SEP endopept
28	16.2	77.1	2580	24	ABN84280	Human neprilysin-1
29	16.2	77.1	2636	22	AAF59660	Human neprilysin-1
30	16.2	77.1	2663	22	AAF59661	CDNA encoding neut
31	16.2	77.1	2676	21	AAK63764	Human neprilysin-1
32	16.2	77.1	2714	22	AAF59659	Human SEP endopept
33	16.2	77.1	2893	24	ABN84279	Human soluble secr
34	16.2	77.1	2953	24	ABK48251	CDNA encoding nove
35	16.2	77.1	2975	24	AAD28547	Human SEP cDNA inc
36	16.2	77.1	5610	23	ABL03669	Drosophila melanog
37	16.2	77.1	5610	23	ABL03669	Drosophila melanog
C 38	16.2	77.1	6162	23	ABL22746	Drosophila melanog
C 39	16.2	77.1	6932	23	ABL03668	Drosophila melanog
40	16.2	77.1	9063	23	ABL03668	Mouse laminin alph
41	16.2	77.1	11009	24	ABQ72907	Mouse laminin-15 a
42	16.2	77.1	11009	24	AAI70816	Human prostate exp
43	15.8	75.2	284	23	ABV38630	Human secretory pr
44	15.8	75.2	411	22	AAC90709	Human cDNA 5'-end
45	15.8	75.2	690	22	AAK92287	

ALIGNMENTS

RESULT 1  
AAZ28815  
ID AAZ28815 standard; DNA; 21 BP.  
XX AAZ28815;  
XX AAZ28815;  
XX 01-FEB-2000 (first entry)  
DE Rat membrane metalloprotease NEPII gene probe #5.

Rat: membrane metalloprotease; neprilysin II; NEPII; inactivation; ss;  
neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.

OS Synthetic.  
OS Rattus rattus.

XX FR2777291-A1.

XX 15-OCT-1999.

XX 08-APR-1998; 98PR-0004389.

XX 08-APR-1998; 98PR-0004389.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;

PI Schwartz JC;

DR WPI; 1999-593429/51.

male erectile dysfunction (MED) or female sexual dysfunction (FSD) (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are also useful for treating the above disorders and other disorders such as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and hypoactive sexual desire disorder. The present sequence is SEP consensus DNA sequence found in human, mouse and rat.

```

XX SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;
Query Match          92.4%; Score 19.4; DB 24; Length 2286;
Best Local Similarity 95.2%; Pred. No. 3.6;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTACCCCAAGCTGCGTGATAG 21
Db 246 CCACCCCAAGCTGCGTGATAG 266

RESULT 4
AAC48434/G.
ID AAC48434 standard; DNA; 1177 BP.
XX AC AAC48434;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 57464.
XX KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX .PR
PR 05-MAR-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123380.
PR 23-MAR-1999; 99US-0123548.
PR 29-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144083.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.

```

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157723.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160983.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match: 84.8%; Score 17.8; DB 21; Length 1177;  
Best Local Similarity 90.5%; Pred. No. 22;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGGTGATAG 21  
Db 271 CAACACCAAGCTGGTGATAG 251  
RESULT 5  
AAC39055/c  
ID AAC39055 standard; DNA; 1218 BP.  
XX AC AAC39055;  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23200.  
XX Arabidopsis thaliana.  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142970.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 84.8%; Score 17.8; DB 21; Length 1218;  
Best Local Similarity 90.5%; Pred No. 22;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 CTACCCCAAGTCGCTGATAG 21  
- - - - -  
Db 282 CAACACCAAGTCGCTGATAG 262

RESULT 6  
AAA63763

ID AAAG3763 standard; cDNA; 2925 BP.  
 AC AAAG3763;  
 XX  
 DT 04-DEC-2000 (first entry)  
 DE  
 XX  
 XX cDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.  
 DE Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
 XX NEP-like enzyme; protein production; protein secretion;  
 KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
 KW fertility; bone disease; abnormal phosphate metabolism; ss.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 AC 332..2629  
 FT CDS /\*tag= a  
 FT /product= "neutral endopeptidase metalloproteinase-like  
 FT enzyme NL-1"  
 FT  
 FT  
 PN WO2000047750-A2.  
 XX  
 XX 17-AUG-2000.  
 PD  
 XX 11-FEB-2000; 2000WO-CA00147.  
 PF  
 XX 11-FEB-1999; 99CA-2260376.  
 PR  
 XX (UYMO-) UNIV MONTREAL.  
 PA  
 XX Desrozeillers L, Boileau G;  
 PI  
 XX WPI; 2000-549148/50.  
 DR P-PSDB; AAB08130.  
 DR  
 XX Novel neutral endopeptidase-like metalloproteinase polypeptides and  
 PT polynucleotides, used to screen for related sequences and enzyme  
 PT inhibitors, used for the treatment of NL-3 related bone disorders -  
 PS Disclosure; Fig 3; 59pp; English.  
 XX  
 CC The present sequence encodes a murine neutral endopeptidase  
 CC metalloproteinase-like enzyme, designated NL-1. The specification  
 CC also describes NL-2 and NL-3. The NL enzymes are used to test for  
 CC specific inhibitors. The N-terminal region of the enzymes can be used  
 CC to promote production and secretion of foreign proteins and active  
 CC biopeptides, using chimeric constructs containing the foreign protein  
 CC downstream from and in phase with the N-terminal region. The NL enzymes  
 CC are have been localised to the brain, and may be useful in the  
 CC treatment of neurological diseases such as Alzheimer's disease, pain,  
 CC and psychiatric disorders. NL enzymes have also been localised to the  
 CC testis and ovaries, and may be used to control fertility. They have  
 CC also been localised to bones, and may be used to treat bone diseases,  
 CC and abnormal phosphate metabolisms related to improper peptide  
 CC processing by the NL-3 enzyme.  
 XX  
 SQ Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;  
 Query Match 84.8%; Score 17.8; DB 21; Length 2925;  
 Best Local Similarity 90.5%; Pred. No. 24;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CTACCCCAAGCTGCTGTATG 21  
 | ||||| ||||| |||||  
 DB 558 CCACCCCAAGCTGCTGTATG 578  
 RESULT 7  
 ABO55050  
 ID ABO55050 standard; cDNA; 1521 BP.  
 XX  
 AC ABO55050;  
 XX  
 DT 22-AUG-2002 (first entry)  
 DE Human ovarian antigen HAOTXG2 cDNA, SEQ ID NO:930.  
 XX  
 XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW gastrointestinal disorder; respiratory disorder; neurological disorder;  
 KW cardiovascular disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200200677-A1.  
 PN  
 XX 03-JAN-2002.  
 PD  
 XX 07-JUN-2001; 2001WO-US18569.  
 PF  
 XX 07-JUN-2000; 2000US-209467P.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Birse CE, Rosen CA;  
 PI  
 XX WPI; 2002-147878/19.  
 DR P-PSDB; ABO41973.  
 DR  
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
 PT ovarian cancer), immune disorders, cardiovascular disorders and  
 PT neurological diseases -  
 XX  
 PS Claim 1; SEQ ID NO 930; 2922pp; English.  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention.  
 CC  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1521 BP; 436 A; 321 C; 336 G; 421 T; 7 other;

Query Match 80.0%; Score 16.8; DB 24; Length 1521;  
 Best Local Similarity 90.0%; Pred. No. 72;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGGTGATA 20  
 |||||  
 Db 361 CTACCCCAAGCTGGTGAATA 380

RESULT 8  
 ABK35251  
 ID ABK35251 standard; cDNA; 2026 BP.  
 XX AC ABK35251;  
 XX DT 08-MAY-2002 (first entry)  
 XX DE Human cDNA encoding secreted protein #389.  
 XX KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;  
 KW viral infection; bacterial infection; fungal infection; diabetes; asthma;  
 KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disease;  
 KW coagulation disorder; inflammatory disorder; Crohn's disease; cancer; ulcer;  
 KW tissue regeneration; wound healing; burn; haematopoesis;  
 KW myeloid cell deficiency; lymphoid cell deficiency.  
 XX KW Homo sapiens.  
 OS WO200177288-A2.  
 XX PN 18-OCT-2001.  
 XX PD 29-MAR-2001; 2001WO-US10224.  
 XX PF 06-APR-2000; 2000US-195582P.  
 XX PR (GEMY ) GENETICS INST INC.  
 XX PA Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
 XX PI Gulukota K, Graham JR;  
 XX PI WPI; 2002-179321/23.

Five hundred and ninety two polynucleotides derived from a variety of human tissue sources which encode secreted proteins, useful for treating immune deficiencies and disorders such as autoimmune disorders

Claim 1; Page 274-275; 372pp; English.

The invention relates to 592 polynucleotides which have been derived from a variety of human tissue sources and which encode novel secreted proteins. The polynucleotides can be used as probes for the identification and isolation of full length cDNA and genomic DNA. The polynucleotides and proteins can also be used as nutritional supplements. The proteins are useful in the treatment of various immune deficiencies and disorders such as viral infections, bacterial infections, fungal infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions and conditions (e.g. asthma). They are also useful for treating neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's disease), liver fibrosis, coagulation disorders (e.g. haemophilia), inflammatory disorders (e.g. Crohn's disease) and tumours. They are also useful for tissue regeneration, for wound healing and in the treatment of burns, incisions and ulcers. The proteins are also useful for regulating haematopoesis and for treating myeloid or lymphoid cell deficiencies. Sequences ABK34863-ABK35454 represent polynucleotides of the invention.

Sequence 2026 BP; 624 A; 426 C; 425 G; 551 T; 0 other;

Query Match 80.0%; Score 16.8; DB 24; Length 2026;

Best Local Similarity 90.0%; Pred. No. 74;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGGTGATA 20  
 |||||  
 Db 1501 CTACCCCAAGCTGGTGAATA 1520

RESULT 9  
 AAV58755  
 ID AAV58755 standard; cDNA; 2835 BP.  
 XX AC AAV58755;  
 XX DT 18-JAN-1999 (first entry)  
 XX DE Human secreted protein bp140\_1 cDNA.  
 XX KW Secreted protein; human; bp140\_1; ds.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 XX CDS 61..867 /\*tag= a  
 XX FT 641..1651  
 XX FT /\*tag= b  
 XX PN WO9840404-A2.  
 XX PD 17-SEP-1998.  
 XX PF 09-MAR-1998; 98WO-US04601.  
 XX PR 06-MAR-1998; 98US-0036321.  
 XX PR 11-MAR-1997; 97US-0815381.  
 XX PA (GEMY ) GENETICS INST INC.  
 XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;  
 XX PI Racie LA, Spaulding V, Treacy M;  
 XX XX WPI; 1998-520802/44.  
 XX DR P-PSDB; AAW69424-25.

New isolated polynucleotides and secreted proteins - useful as, e.g. nutritional additives, immunostimulators, haematopoesis regulators and as diagnostic agents

Claim 4; Page 68-69; 119pp; English.  
 This full-length cDNA clone, designated bg140.1, was isolated from a human adult brain cDNA library using methods which are selective for cDNAs encoding secreted proteins, or was identified as encoding a secreted or transmembrane protein on the basis of computer analysis of the encoding protein. It includes alternative reading frames encoding bg140.1 protein (see AAW69424-25). Homology is shown to some database sequences. The invention provides isolated polynucleotides (see AAV58754-63) obtained from human adult testis, brain, retina or placenta, or from foetal kidney or brain cDNA libraries. These are all deposited as ATCC 98353. They encode novel human secreted proteins (see AAW69423-33) that may have e.g. nutritional activity, immune stimulating or suppressing activity, haematopoesis regulating activity, tissue growth activity, activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, antiinflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition or other activities. They may also be used for diagnostic purposes.

Sequence 2835 BP; 869 A; 590 C; 622 G; 753 T; 1 other;

Query Match 80.0%; Score 16.8; DB 19; Length 2835;

Best Local Similarity 90.0%; Pred. No. 77;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20  
|||||  
Db 1711 CTACCCCAAGCTGCGTGATA 1730

RESULT 10  
ABK84434  
ID ABR84434 standard; cDNA; 3328 BP.  
XX  
AC ABR84434;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #1005.  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200228999-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001WO-US30821.  
XX  
PR 03-OCT-2000; 2000US-237189P.  
XX  
PA (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX  
DR WPI; 2002-435328/46.  
XX  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity  
XX  
PS Claim 1: SEQ ID No 1005; 114pp; English.  
XX  
CC The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis, as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GCA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GCA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic

response in a subject, exposure of a subject to a pathogen or sterile  
inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
inflammatory bowel disease, ARDS, adult respiratory distress syndrome,  
periodontal disease; also bacterial infection, fungal infection, and M5 is  
parasitic infection, protozoal infection, viral infection, and M5 is  
useful for treating one of the above conditions. The present  
sequence represents a gene differentially expressed in granulocytes.  
Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 3328 BP; 1052 A; 676 C; 768 G; 832 T; 0 other;

Query Match 80.0%; Score 16.8; DB 24; Length 3328;  
Best Local Similarity 90.0%; Pred. No. 78;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20  
|||||  
Db 2940 CTACCCCAAGCTGCGTGATA 2959

RESULT 11  
AAS84900  
ID AAS84900 standard; cDNA; 3610 BP.  
XX  
AC AAS84900;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20704.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR P-PSDB; ABG20713.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1: SEQ ID No 20704; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: the sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3610 BP; 1009 A; 842 C; 848 G; 909 T; 2 other;

Query Match 80.0%; Score 16.8; DB 23; Length 3610;

Best Local Similarity 90.0%; Pred. No. 79;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20

||||| ||||| || |||

Db 2500 CTACCCCAAGCTGCGTGATA 2519

RESULT 12

AAS84899

ID AAS84899 standard; cDNA; 4424 BP.

XX AAS84899;

DT 13-FEB-2002 (first entry)

DE DNA encoding: novel human diagnostic protein #20703.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG20712.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX Claim 1; SEQ ID No 20703; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: the sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 4424 BP; 1282 A; 1001 C; 1041 G; 1100 T; 0 other;

Query Match 80.0%; Score 16.8; DB 23; Length 4424;

Best Local Similarity 90.0%; Pred. No. 81;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20

||||| ||||| || |||

Db 3314 CTACCCCAAGCTGCGTGATA 3333

RESULT 13

AAS28366

ID AAS28366 standard; DNA; 7606 BP.

XX AAS28366;

DT 07-NOV-2001 (first entry)

DE Genomic sequence #206 encoding for novel human respiratory antigen.

XX Human; respiratory antigen; respiratory disorder; throat disorder;  
 KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;  
 KW anti allergic; anti asthmatic; anti inflammatory; olfactory;  
 KW respiratory active; ds.

XX Homo sapiens.

XX WO200155448-A1.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01333.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUL-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226581.  
PR 22-AUG-2000; 2000US-0226588.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234277.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI. INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-476224/51.  
XX  
XX Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the respiratory system including respiratory  
PT cancers and also for testing and detection e.g. diagnosis -  
XX  
PS Disclosure; SED ID No 800; 546pp; English.  
XX  
XX The present invention relates to the isolation of novel human  
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic  
CC sequences encoding for these polypeptides. The sequences of the  
CC invention are useful for preventing, treating and/or prognosing  
CC disorders related to the respiratory system including throat  
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),  
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,  
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of  
CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences  
CC of the invention are useful in gene therapy and antisense therapy.  
CC AAS28161-AAS28764 represent genomic sequences encoding for novel  
CC human respiratory antigens.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 7606 BP; 2043 A; 1602 C; 1685 G; 2276 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 7606;  
Best Local Similarity 90.0%; Pred. No. 86;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTACCCCAAGCTCGTGATA 20  
|||||  
Db 6503 CTACCCCAAGCTCGTGATA 6522

RESULT 14  
AAS28367  
ID AAS28367 standard; DNA; 7608 BP.  
XX AC AAS28367;  
XX DT 07-NOV-2001 (first entry)  
XX DE Genomic sequence #207 encoding for novel human respiratory antigen.  
XX KW Human; respiratory antigen; respiratory disorder; throat disorder;  
XX KW lung disorder; nose disorder; lung cancer; gene therapy; cytostatic;  
XX KW anti allergic; anti asthmatic; anti inflammatory; olfactory;  
XX KW respiratory active; ds.  
OS Homo sapiens.  
XX PN WO200155448-A1.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01333.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184564.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225113.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 12-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.

```

17-NOV-2000; 2000US-0249300.
01-DEC-2000; 2000US-0250160.
01-DEC-2000; 2000US-0250391.
05-DEC-2000; 2000US-0251030.
05-DEC-2000; 2000US-0251988.
05-DEC-2000; 2000US-0256719.
06-DEC-2000; 2000US-0251479.
08-DEC-2000; 2000US-0251856.
08-DEC-2000; 2000US-0251868.
08-DEC-2000; 2000US-0251869.
08-DEC-2000; 2000US-0251989.
08-DEC-2000; 2000US-0251990.
11-DEC-2000; 2000US-0254097.
05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;
WPI; 2001-476224/51.

Isolated polypeptide for treating, preventing and/or prognosing
disorders related to the respiratory system including respiratory
cancers and also for testing and detection e.g. diagnosis -
Disclosure; SED ID No 801; 546pp; English.

The present invention relates to the isolation of novel human
respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
sequences encoding for these polypeptides. The sequences of the
invention are useful for preventing, treating and/or prognosing
disorders related to the respiratory system including throat
disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
the respiratory tissues e.g. lung cancer. The polynucleotide sequences
of the invention are useful in gene therapy and antisense therapy.
AAS281611-AAS28764 represent genomic sequences encoding for novel
human respiratory antigens.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO.
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 7608 BP; 2043 A; 1599 C; 1690 G; 2276 T; 0 other;

Query Match 80.0%; Score 16.8; DB 22; Length 7608;
Best Local Similarity 90.0%; Pred.No. 86;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20
DB 11111111111111111111
6505 CTACCCCAAGCTGTGTAATA 6524

RESULT 15
AAL37558
ID AAL37558 standard; DNA; 12771 BP.
XX AC
XX AAL37558;
XX DT
XX DT 08-JAN-2002 (first entry)
XX DE
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 3923.
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antididiabetic; antiinflammatory; antiulcer;
XX KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein;
XX KW musculoskeletal system; ds.
XX OS
XX Homo sapiens.

```

PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HOMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-451937/48.

XX Isolated polypeptide for treating, preventing and/ or prognosing  
PT disorders related to the musculoskeletal system including  
PT musculoskeletal cancers and also for testing and detection e.g.  
PT diagnosis -

XX Example 2; SEQ ID NO 3923; 781pp + Sequence Listing; English.

PS The invention relates to novel genes (AAL34669-AAL37666) and proteins  
CC (ABB03087-ABB04109) associated with the musculoskeletal system useful  
CC for preventing, treating or ameliorating medical conditions e.g. by  
CC protein or gene therapy. The genes are isolated from a range of human  
CC tissues disclosed in the specification. The nucleic acids, proteins,  
CC antibodies and (ant)agonists are useful in the diagnosis, treatment  
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and  
CC other cancers of the adrenal gland, bone, bone marrow, breast,  
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;  
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound  
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;  
CC and (f) infectious diseases such as viral, bacterial, fungal and  
CC parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 12771 BP; 3627 A; 2428 C; 2622 G; 4094 T; 0 other;

Query Match

Best Local Similarity 80.0%; Score 16.8; DB 22; Length 12771;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCCCAAGCTGCGTGATAG 21

||||||| ||| |||||

Db 6219 TACCCCAACCTGGTGATAG 6238

Search completed: July 8, 2003, 02:18:55

Job time : 132.941 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 28.9902 Seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-9

Perfect score: 21

Sequence: 1 ctaccccaagctgctgatag 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/1/ina/3A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/3B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.4	78.1	629	2	US-08-698-805-3
2	15.8	75.2	1482	4	US-09-134-001C-1551
3	15.4	73.3	4403765	4	US-09-103-840A-2
4	15.4	73.3	4411529	4	US-09-103-840A-1
5	14.8	70.5	849	1	US-08-667-023-1
6	14.8	70.5	2439	4	US-09-233-989-1
7	14.8	70.5	2443	1	US-08-452-262-1
8	14.8	70.5	2443	1	US-08-734-550-1
9	14.8	70.5	2443	5	PCT-US96-07528-1
10	14.8	70.5	3117	3	US-08-909-954-3
11	14.6	69.5	1674	1	US-08-523-855A-1
12	14.6	69.5	3226	3	US-08-539-208A-5
13	14.6	69.5	152331	3	US-09-128-155-16
14	14.6	69.5	176373	3	US-09-128-155-17
15	14.4	68.6	4403765	4	US-09-103-840A-2
16	14.4	68.6	4411529	4	US-09-103-840A-1
17	14.2	67.6	510	3	US-08-441-971-66
18	14.2	67.6	510	4	US-08-221-653-66
19	14.2	67.6	510	4	US-08-442-144A-66
20	14.2	67.6	510	4	US-08-441-970-66
21	14.2	67.6	999	2	US-08-924-759-3
22	14.2	67.6	999	3	US-09-248-335-3
23	14.2	67.6	1271	4	US-09-231-227-1
24	14.2	67.6	1808	1	US-08-351-149-4
25	14.2	67.6	1808	3	US-08-384-828-4
26	14.2	67.6	1808	3	US-08-895-474-4
27	14.2	67.6	2682	1	US-07-971-819A-1

28 14.2 67.6 2682 1 US-07-977-434-3  
29 14.2 67.6 2682 1 US-08-475-231-1  
30 14.2 67.6 2682 1 US-08-458-819-3  
31 14.2 67.6 2682 4 US-09-105-697-9  
32 14.2 67.6 2682 5 PCT-US91-07035-3  
33 14.2 67.6 8133 4 US-09-659-791A-10  
34 14.2 67.6 9997 1 US-08-246-982A-15  
35 14.2 67.6 9997 1 US-08-453-265-15  
36 14.2 67.6 10103 2 US-08-457-273B-7  
37 14.2 67.6 15297 4 US-09-817-180-3  
38 14.2 67.6 50341 1 US-08-247-901C-1  
39 14.2 67.6 50341 2 US-09-075-904-1  
40 14.2 67.6 52297 4 US-09-426-436-1  
41 14.2 67.6 52297 4 US-08-705-557-1  
42 13.8 65.7 424 4 US-09-641-638-243  
43 13.8 65.7 424 4 US-09-641-638-244  
44 13.8 65.7 424 4 US-09-641-638-245  
45 13.8 65.7 483 4 US-09-392-184-21

## ALIGNMENTS

RESULT 1  
US-08-698-805-3  
; Sequence 3, Application US/08698805  
; Patent No. 5869288  
; GENERAL INFORMATION:  
; APPLICANT: Chapman, Martin  
; APPLICANT: Artuda, L. Karla  
; TITLE OF INVENTION: Molecular Cloning of Cochroach  
; TITLE OF INVENTION: Allergens, Amino Acid and Nucleotide Sequences Thereof  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Obolon, Spivak, McClelland, Maier & Neustadt,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/698,805  
; APPLICATION NUMBER: US/08/698,805  
; FILING DATE: 16-AUG-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/002,510  
; FILING DATE: 18-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kelber, Steven B.  
; REGISTRATION NUMBER: 30,073  
; REFERENCE/DOCKET NUMBER: 494-203-27  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-413-3000  
; TELEFAX: 703-413-2220  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 629 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 2..547  
; US-08-698-805-3

Sequence 3, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 9, Appli  
Sequence 3, Appli  
Sequence 10, Appli  
Sequence 15, Appli  
Sequence 7, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 243, App  
Sequence 244, App  
Sequence 245, App  
Sequence 21, Appl

Query Match 78.1%; Score 16.4; DB 2; Length 629;  
Best Local Similarity 94.4%; Pred. No. 14;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ACCCAAGCTGGTGATG 20  
|||||  
Db 417 ACCCAAGCTGGTGATG 434

## RESULT 2

US-09-134-001C-1551/c  
; Sequence 1551, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: LYNN DOUCETTE-STAMM ET AL  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1551  
; LENGTH: 1482  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1551

Query Match 75.2%; Score 15.8; DB 4; Length 1482;  
Best Local Similarity 89.5%; Pred. No. 30;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ACCCAAGCTGGTGATG 21  
|||||  
Db 1401 ACCCAAGCTGGTGATG 1383

## RESULT 3

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 73.3%; Score 15.4; DB 4; Length 4403765;  
Best Local Similarity 94.1%; Pred. No. 53;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGGTG 17  
|||||  
Db 2446460 CTACCCCAAGCTCGGTG 2446476

RESULT 4  
US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 73.3%; Score 15.4; DB 4; Length 4411529;  
Best Local Similarity 94.1%; Pred. No. 53;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGGTG 17  
|||||  
Db 2449159 CTACCCCAAGCTCGGTG 2449175

## RESULT 5

US-08-667-023-1  
; Sequence 1, Application US/08667023  
; Patent No. 5817783  
; GENERAL INFORMATION:  
; APPLICANT: Callabretta, Bruno  
; APPLICANT: Venturelli, Donatella  
; APPLICANT: Martinez, Robert V.  
; TITLE OF INVENTION: DR-0023 AND COMPOSITIONS, METHODS OF MAKING AND  
; TITLE OF INVENTION: METHODS OF USING THE SAME  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5817783ris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/667,023  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/000,427  
; FILING DATE: 22-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Deluca, Mark  
; REGISTRATION NUMBER: 33,229  
; REFERENCE/DOCKET NUMBER: TJU-1992  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 849 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 19..525  
US-08-667-023-1

Query Match 70.5%; Score 14.8; DB 1; Length 849;  
Best Local Similarity 88.9%; Pred. No. 93;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGA 18  
||||| || ||||| |||||  
Db 222 CTACCGCGAGCTGCGTGA 239

## RESULT 6

US-09-233-989-1  
Sequence 1, Application US/09233989  
Patent No. 6248527

GENERAL INFORMATION:

APPLICANT: Chen, Hong

APPLICANT: Meyer, Joanne

TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on

FILE REFERENCE: Mutations Found in Carboxypeptidase E

CURRENT APPLICATION NUMBER: US/09/233,989

CURRENT FILING DATE: 1999-01-19

EARLIER APPLICATION NUMBER: 60/105,102

EARLIER FILING DATE: 1998-10-21

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO.1

LENGTH: 2439

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (287)..(1714)

FEATURE:

OTHER INFORMATION: carboxypeptidase E

US-09-233-989-1

Query Match 70.5%; Score 14.8; DB 4; Length 2439;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGA 18  
||||| ||||| |||||  
Db 451 CTACCGCGAGCTGCGCA 468

## RESULT 7

US-08-452-262-1

Sequence 1, Application US/08452262

Patent No. 5593837

GENERAL INFORMATION:

APPLICANT: Nagert, Jurgen K.

APPLICANT: Leiter, Edward H.

TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH

TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kevin M. Farrell, P.C.

STREET: P.O. Box 999

CITY: York Harbor

STATE: ME

COUNTRY: US

ZIP: 03911

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,262  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Farrell, Kevin M.  
REGISTRATION NUMBER: 35,505  
REFERENCE/DOCKET NUMBER: JL-9501  
TELEPHONE: (207) 363-0558  
TELEFAX: (207) 363-0528  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2443 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 414..1721  
US-08-452-262-1

Query Match 70.5%; Score 14.8; DB 1; Length 2443;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGA 18  
||||| ||||| |||||  
Db 455 CTACCGCGAGCTGCGCA 472

## RESULT 8

US-08-734-550-1

Sequence 1, Application US/08734550

Patent No. 5650932

GENERAL INFORMATION:

APPLICANT: Nagert, Jurgen K.

APPLICANT: Leiter, Edward H.

TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH

TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kevin M. Farrell, P.C.

STREET: P.O. Box 999

CITY: York Harbor

STATE: ME

COUNTRY: US

ZIP: 03911

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/734,550

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Farrell, Kevin M.

REGISTRATION NUMBER: 35,505

REFERENCE/DOCKET NUMBER: JL-9501

TELEPHONE: (207) 363-0558

TELEFAX: (207) 363-0528

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2443 base pairs

;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: mrna  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 414..1721  
US-08-734-550-1

Query Match 70.5%; Score 14.8; DB 1; Length 2443;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGA 18  
||||||| ||||||| ||  
Db 455 CTACCCGAGCTGCGGA 472

## RESULT 9

PCT-US96-07528-1

; Sequence 1, Application PC/TUS9607528

; GENERAL INFORMATION:

; APPLICANT: The Jackson Laboratory

; TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH

; TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kevin M. Farrell, P.C.

; STREET: P.O. Box 999

; CITY: York Harbor

; STATE: ME

; COUNTRY: US

; ZIP: 03911

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US96/07528

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/452,262

; FILING DATE: 26-MAY-1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Farrell, Kevin M.

; REGISTRATION NUMBER: 35,505

; REFERENCE/DOCKET NUMBER: JL-9501 WO

; TELEPHONE: (207) 363-0558

; TELEFAX: (207) 363-0528

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2443 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cdna

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 414..1721

PCT-US96-07528-1

Query Match 70.5%; Score 14.8; DB 5; Length 2443;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGA 18  
||||||| ||||||| ||  
Db 455 CTACCCGAGCTGCGGA 472

## RESULT 10

US-08-909-954-3

; Sequence 3, Application US/08909954A

; Patent No. 6100058

; GENERAL INFORMATION:

; APPLICANT: Allen, Maxine J.

; APPLICANT: Buckler, Alan J.

; TITLE OF INVENTION: GAP12 Genes and their uses

; FILE REFERENCE: SEQ-11P

; CURRENT APPLICATION NUMBER: US/08/909,954A

; CURRENT FILING DATE: 1997-08-12

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO: 3

; LENGTH: 3117

; TYPE: DNA

; ORGANISM: M. musculus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (0)...(0)

US-08-909-954-3

Query Match 70.5%; Score 14.8; DB 3; Length 3117;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ACCCCCAAGCTGCGTGATA 20  
||||||| |||||  
Db 2820 ACCCCCAAGCTGCGTGATA 2837

## RESULT 11

US-08-523-855A-1

; Sequence 1, Application US/08523855A

; Patent No. 5824538

; GENERAL INFORMATION:

; APPLICANT: Arthur A. Branstrom

; APPLICANT: Donata R. Sizemore

; APPLICANT: Jerald C. Sadoff

; TITLE OF INVENTION: Bacterial Delivery System

; NUMBER OF SEQUENCES: 8

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: John Moran

; STREET: USA MRMC - MCMR-JA

; CITY: FORT DETRICK, FREDERICK

; STATE: MARYLAND

; COUNTRY: USA

; ZIP: 21702-5012

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Macintosh 7.5

; SOFTWARE: Microsoft Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/523,855A

; FILING DATE: 09/06/95

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Moran, John

; REGISTRATION NUMBER: 26,313

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 619-2065

; TELEFAX: (301) 619-7714

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1674 base pairs

; TYPE: Nucleic acid

; STRANDEDNESS: double

```
; TOPOLOGY: Linear
; US-08-523-855A-1
;
; Query Match
; Best Local Similarity 69.5%; Score 14.6; DB 1; Length 1674;
; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 CTACCCCAAGCTGCGTGATAG 21
;      ||||| ||||| ||||| ||
; Db 482 CTATCCAAAGCTGCGTGAG 502
;
;
; RESULT 12
; US-08-539-205A-5/C
; Sequence 5, Application US/08539205A
; Patent No. 6001619
; GENERAL INFORMATION:
; APPLICANT: Beach, David H.
; APPLICANT: Caligiuri, Maureen
; APPLICANT: Nefsky, Bradley
; TITLE OF INVENTION: Ubiquitin Ligases, and Uses Related Thereto
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/539,205A
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: CSV-005.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 832-1000
; TELEFAX: (617) 832-7000
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3226 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 400..2901
; US-08-539-205A-5
;
; Query Match
; Best Local Similarity 69.5%; Score 14.6; DB 3; Length 3226;
; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 CTACCCCAAGCTGCGTGATAG 21
;      ||||| ||||| ||||| ||
; Db 1243 CTCTGCAAGCTGCGTGATAG 1223
;
;
; RESULT 13
; US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; AND USES THEREOF
;
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-16
;
; Query Match
; Best Local Similarity 69.5%; Score 14.6; DB 3; Length 152331;
; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 CTACCCCAAGCTGCGTGATAG 21
;      ||||| ||||| ||||| ||
; Db 14072 CTACCCCAAGCTGCGTGACAG 14092
;
;
; RESULT 14
; US-09-128-155-17
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
; US-09-128-155-17
;
; Query Match
; Best Local Similarity 69.5%; Score 14.6; DB 3; Length 176373;
; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 CTACCCCAAGCTGCGTGATAG 21
;      ||||| ||||| ||||| ||
; Db 103933 CTACCCCAAGCTGCGTGACAG 103953
;
;
; RESULT 15
; US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
```

;; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
;; TITLE OF INVENTION: TUBERCULOSIS  
;; FILE REFERENCE: 24366-20007.00  
;; CURRENT APPLICATION NUMBER: US/09/103,840A  
;; CURRENT FILING DATE: 1998-06-24  
;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 2  
;; LENGTH: 4403765  
;; TYPE: DNA  
;; ORGANISM: Mycobacterium tuberculosis  
;; FEATURE:  
;; OTHER INFORMATION: CDC 1551  
;; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
;; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 68.6%; Score 14.4; DB 4; Length 4403765;  
Best Local Similarity 93.8%; Pred. No. 69;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 3 ACCCCAGCTGCGTGA 18  
||||| |||||||  
Db: 2617028 ACCCCAGCTGCGTGA 2617013

Search completed: July 8, 2003, 09:31:28  
Job time : 49.0402 secs

Query Match	80.0%;	Score 16.8;	DB 9;	Length 364;
Best Local Similarity	90.0%;	Pred. No. 25;		
Matches	18;	Conservative 0;	Mismatches 2;	Indels 0;
QY	1	CTACCCCAAGCTGCGTGATA	20	
Db	145	CTACCCCAAGCTGTGTAATA	126	

## RESULT 2

US-09-918-995-34494  
; Sequence 34494, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 34494  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-34494

Query Match 80.0%; Score 16.8; DB 9; Length 401;  
Best Local Similarity 90.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20  
Db 103 CTACCCCAAGCTGTGTAATA 122  
|||||

## RESULT 3

US-10-084-817-324  
; Sequence 324, Application US/10084817  
; Publication No. US20030119009A1  
; GENERAL INFORMATION:  
; APPLICANT: Susan Stuart  
; APPLICANT: Jed G. Nuchtern  
; APPLICANT: Sharon E. Plon  
; APPLICANT: Jason M. Shohet  
; TITLE OF INVENTION: GENES REGULATED BY MYCIN ACTIVATION  
; FILE REFERENCE: PA-0046 US  
; CURRENT APPLICATION NUMBER: US/10/084,817  
; CURRENT FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: 60/270,784  
; PRIOR FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 365  
; SOFTWARE: PERL Program  
; SEQ ID NO 324  
; LENGTH: 1316  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. US20030119009A1 988665.10  
; LOCATION: 793, 798  
; OTHER INFORMATION: a, t, c, g, or other  
US-10-084-817-324

Query Match 80.0%; Score 16.8; DB 9; Length 1316;  
Best Local Similarity 90.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20  
Db 425 CTACCCCAAGCTGTGTAATA 444  
|||||

## RESULT 4

US-09-822-849A-389  
; Sequence 389, Application US/09822849A  
; Patent No. US20020045170A1  
; GENERAL INFORMATION:

; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakari  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6403  
; CURRENT APPLICATION NUMBER: US/09/822,849A  
; CURRENT FILING DATE: 2001-09-04  
; PRIOR APPLICATION NUMBER: 60/195,582  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 598  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 389  
; LENGTH: 2026  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-849A-389

Query Match 80.0%; Score 16.8; DB 10; Length 2026;  
Best Local Similarity 90.0%; Pred. No. 25;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20  
Db 1501 CTACCCCAAGCTGTGTAATA 1520  
|||||

## RESULT 5

US-10-074-095-800  
; Sequence 800, Application US/10074095  
; Publication No. US20030077704A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC008C1  
; CURRENT APPLICATION NUMBER: US/10/074,095  
; CURRENT FILING DATE: 2002-02-14  
; PRIOR APPLICATION NUMBER: 09/764,860  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/225,757  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/226,868  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/216,647  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,267  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/216,880  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: 60/225,270

PRIOR APPLICATION NUMBER: 60/241,787	PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474	PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532	PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681	PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182	PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,211	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836	PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/230,438	PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/215,135	PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 60/225,266	PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/249,218	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,208	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,213	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,212	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,207	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,245	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,244	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,217	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,211	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,215	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,264	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,214	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,297	PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/232,400	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/231,242	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,081	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/232,080	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,414	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/231,244	PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/233,064	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/233,063	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,397	PRIOR FILING DATE: 2000-09-14
PRIOR APPLICATION NUMBER: 60/232,399	PRIOR FILING DATE: 2000-09-14

4  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842

PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/246,532  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/249,216  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,210  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/226,681  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,759  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/225,213  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/227,182  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/225,214  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/235,836  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/230,438  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/215,135  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 60/225,266  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/249,218  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,208  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,213  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,212  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,207  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,245  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,244  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,217  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,211  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,215  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,264  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/249,214  
PRIOR FILING DATE: 2000-11-17

PRIOR APPLICATION NUMBER: 60/249,297  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/232,400  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/231,242  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,081  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/232,080  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,414  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/231,244  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/233,064  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/233,063  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,397  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,399  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/232,401  
PRIOR FILING DATE: 2000-09-14  
PRIOR APPLICATION NUMBER: 60/241,808  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,826  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,786  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/241,221  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,475  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 60/231,243  
PRIOR FILING DATE: 2000-09-08

Query Match 80.0%; Score 16.8; DB 9; Length 7608;  
Best Local Similarity 90.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTACCCCAAGCTGGTGATA 20  
|||||  
Db 6505 CTACCCCAAGCTGGTGATA 6524

## RESULT 8

US-09-764-860-801  
Sequence 801, Application US/09764860  
Patent No. US20020094953A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC008  
CURRENT APPLICATION NUMBER: US/09764,860  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1198  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 801  
LENGTH: 7608  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-860-801

Query Match 80.0%; Score 16.8; DB 10; Length 7608;  
Best Local Similarity 90.0%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CTACCCCAAGCTGGTGATA 20  
|||||  
Db 6505 CTACCCCAAGCTGGTGATA 6524

RESULT 9  
US-09-764-877-3923  
Sequence 3923, Application US/09764877  
Patent No. US20020147140A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC005  
CURRENT APPLICATION NUMBER: US/09/764.877  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 4031  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3923  
LENGTH: 12771  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-877-3923

Query Match 80.0%; Score 16.8; DB 10; Length 12771;  
Best Local Similarity 90.0%; Pred. No. 27;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TACCCCAAGCTGCTGATAG 21  
||||||| ||| |||||  
Db 6219 TACCCCAAGCTGCTGATAG 6238

RESULT 10  
US-10-074-095-798  
Sequence 798, Application US/10074095  
Publication No. US20030077704A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC008C1  
CURRENT APPLICATION NUMBER: US/10/074.095  
CURRENT FILING DATE: 2002-02-14  
PRIOR APPLICATION NUMBER: 09/764.860  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/225,757  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/226,868  
PRIOR FILING DATE: 2000-08-22  
PRIOR APPLICATION NUMBER: 60/216,647  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,267  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/216,880  
PRIOR FILING DATE: 2000-07-07  
PRIOR APPLICATION NUMBER: 60/225,270  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/251,869  
PRIOR FILING DATE: 2000-12-08

PRIOR APPLICATION NUMBER: 60/235,834  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: 60/234,274  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/234,223  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: 60/228,924  
PRIOR FILING DATE: 2000-08-30  
PRIOR APPLICATION NUMBER: 60/224,518  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,369  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/224,519  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,964  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/241,809  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/249,299  
PRIOR FILING DATE: 2000-11-17  
PRIOR APPLICATION NUMBER: 60/236,327  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/241,785  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/244,617  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 60/225,268  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/236,368  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/251,856  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/251,868  
PRIOR FILING DATE: 2000-12-08  
PRIOR APPLICATION NUMBER: 60/229,344  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/234,997  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: 60/229,343  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,345  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,287  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/229,513  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/231,413  
PRIOR FILING DATE: 2000-09-08  
PRIOR APPLICATION NUMBER: 60/229,509  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/236,367  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/237,039  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,038  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/236,370  
PRIOR FILING DATE: 2000-09-29  
PRIOR APPLICATION NUMBER: 60/236,802  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,037  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/237,040  
PRIOR FILING DATE: 2000-10-02  
PRIOR APPLICATION NUMBER: 60/240,960  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/239,935  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/239,937  
PRIOR FILING DATE: 2000-10-13  
PRIOR APPLICATION NUMBER: 60/241,787  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/246,474

;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/246,532  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/249,216  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,210  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/226,681  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,759  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/225,213  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/227,182  
;; PRIOR FILING DATE: 2000-08-22  
;; PRIOR APPLICATION NUMBER: 60/225,214  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/235,836  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: 60/230,438  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/215,135  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: 60/225,266  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/249,218  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,208  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,213  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,212  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,207  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,245  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,244  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,217  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,211  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,215  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,264  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,214  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/249,297  
;; PRIOR FILING DATE: 2000-11-17  
;; PRIOR APPLICATION NUMBER: 60/232,400  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/231,242  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,081  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/232,080  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,414  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/231,244  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,064  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/233,063  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,397  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,399  
;; PRIOR FILING DATE: 2000-09-14  
;; PRIOR APPLICATION NUMBER: 60/232,401  
;; PRIOR FILING DATE: 2000-09-14

;; PRIOR APPLICATION NUMBER: 60/241,808  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,826  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,786  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,221  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,475  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/231,243  
;; PRIOR FILING DATE: 2000-09-08

Query Match. 80.0%; Score 16.8; DB 9; Length 17498;  
Best Local Similarity 90.0%; Pred. No. 27;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTACCCCAAGCTGCGTGATA 20  
|||||  
DB 16395 CTACCCCAAGCTGCGTGATA 16414

RESULT 11  
US-09-764-860-798  
;; Sequence 798, Application US/09764860  
;; Patent No. US20020094953A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
;; FILE REFERENCE: PC008  
;; CURRENT APPLICATION NUMBER: US/09/764,860  
;; CURRENT FILING DATE: 2001-01-17  
;; Prior application data removed - consult PALM or file wrapper  
;; NUMBER OF SEQ ID NOS: 1198  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 798  
;; LENGTH: 17498  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-764-860-798

Query Match 80.0%; Score 16.8; DB 10; Length 17498;  
Best Local Similarity 90.0%; Pred. No. 27;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CTACCCCAAGCTGCGTGATA 20  
|||||  
DB 16395 CTACCCCAAGCTGCGTGATA 16414

RESULT 12  
US-10-074-095-797  
;; Sequence 797, Application US/10074095  
;; Publication No. US2003007704A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
;; FILE REFERENCE: PC008C1  
;; CURRENT APPLICATION NUMBER: US/10/074,095  
;; CURRENT FILING DATE: 2002-02-14  
;; PRIOR APPLICATION NUMBER: 09/764,860  
;; PRIOR FILING DATE: 2001-01-17  
;; PRIOR APPLICATION NUMBER: 60/179,065  
;; PRIOR FILING DATE: 2000-01-31  
;; PRIOR APPLICATION NUMBER: 60/180,628  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: 60/214,886  
;; PRIOR FILING DATE: 2000-06-28  
;; PRIOR APPLICATION NUMBER: 60/217,487  
;; PRIOR FILING DATE: 2000-07-11  
;; PRIOR APPLICATION NUMBER: 60/225,758  
;; PRIOR FILING DATE: 2000-08-14  
;; PRIOR APPLICATION NUMBER: 60/220,963

us-09-647-780a-9.rnpb

wed Jul 9 09:33:58 2003

PRIOR FILING DATE: 2000-07-26  
 PRIOR APPLICATION NUMBER: 60/217,496  
 PRIOR FILING DATE: 2000-07-11  
 PRIOR APPLICATION NUMBER: 60/225,447  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/218,290  
 PRIOR FILING DATE: 2000-07-14  
 PRIOR APPLICATION NUMBER: 60/225,757  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/226,868  
 PRIOR FILING DATE: 2000-08-22  
 PRIOR APPLICATION NUMBER: 60/216,847  
 PRIOR FILING DATE: 2000-07-07  
 PRIOR APPLICATION NUMBER: 60/225,267  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/216,880  
 PRIOR FILING DATE: 2000-07-07  
 PRIOR APPLICATION NUMBER: 60/225,270  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/251,869  
 PRIOR FILING DATE: 2000-12-08  
 PRIOR APPLICATION NUMBER: 60/235,834  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: 60/234,274  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: 60/234,223  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: 60/228,924  
 PRIOR FILING DATE: 2000-08-30  
 PRIOR APPLICATION NUMBER: 60/224,518  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/236,369  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/224,519  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/220,964  
 PRIOR FILING DATE: 2000-07-26  
 PRIOR APPLICATION NUMBER: 60/241,809  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/249,299  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/236,327  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/241,785  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/244,617  
 PRIOR FILING DATE: 2000-11-01  
 PRIOR APPLICATION NUMBER: 60/225,268  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/236,368  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/251,856  
 PRIOR FILING DATE: 2000-12-08  
 PRIOR APPLICATION NUMBER: 60/251,868  
 PRIOR FILING DATE: 2000-12-08  
 PRIOR APPLICATION NUMBER: 60/229,344  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/234,997  
 PRIOR FILING DATE: 2000-09-25  
 PRIOR APPLICATION NUMBER: 60/229,343  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/229,345  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/229,287  
 PRIOR FILING DATE: 2000-09-01  
 PRIOR APPLICATION NUMBER: 60/229,513  
 PRIOR FILING DATE: 2000-09-05  
 PRIOR APPLICATION NUMBER: 60/231,413  
 PRIOR FILING DATE: 2000-09-08  
 PRIOR APPLICATION NUMBER: 60/229,509  
 PRIOR FILING DATE: 2000-09-05  
 PRIOR APPLICATION NUMBER: 60/236,367  
 PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: 60/237,039  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/237,038  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/236,370  
 PRIOR FILING DATE: 2000-09-29  
 PRIOR APPLICATION NUMBER: 60/236,802  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/237,037  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/237,040  
 PRIOR FILING DATE: 2000-10-02  
 PRIOR APPLICATION NUMBER: 60/240,960  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/239,935  
 PRIOR FILING DATE: 2000-10-13  
 PRIOR APPLICATION NUMBER: 60/239,937  
 PRIOR FILING DATE: 2000-10-13  
 PRIOR APPLICATION NUMBER: 60/241,787  
 PRIOR FILING DATE: 2000-10-20  
 PRIOR APPLICATION NUMBER: 60/246,474  
 PRIOR FILING DATE: 2000-11-08  
 PRIOR APPLICATION NUMBER: 60/246,532  
 PRIOR FILING DATE: 2000-11-08  
 PRIOR APPLICATION NUMBER: 60/249,216  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,210  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/226,681  
 PRIOR FILING DATE: 2000-08-22  
 PRIOR APPLICATION NUMBER: 60/225,759  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/225,213  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/227,182  
 PRIOR FILING DATE: 2000-08-22  
 PRIOR APPLICATION NUMBER: 60/225,214  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/235,836  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: 60/230,438  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/215,135  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: 60/225,266  
 PRIOR FILING DATE: 2000-08-14  
 PRIOR APPLICATION NUMBER: 60/249,218  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,208  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,213  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,212  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,207  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,245  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,244  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,217  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,211  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,215  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,264  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,214  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/249,297  
 PRIOR FILING DATE: 2000-11-17  
 PRIOR APPLICATION NUMBER: 60/232,400

RESULT 14  
US-10-017-273A-4

Query Match	77.18;	Score 16.2;	DB 10;	Length 2893;
Best Local Similarity	85.78;	Pred. No. 53;		
Matches 18;	Conservative 0;	Mismatches 3;	Indels 0;	
QY	1	CTACCCCAAGCTGCGTGATAG	21	
Db	453	CCACCCCTGGCTGCGTGATAG	473	

Search completed: July 9, 2003, 02:21:53  
Job time : 122.338 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 231.102 seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-9  
Perfect score: 21  
Sequence: 1 ctaccaccaagctgcgtgatag 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_ov:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_ov:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pin:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vit:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	6	AX014709 Sequence
2	21	100.0	2765	6	AX014701 Sequence
3	21	100.0	174953	2	AC094732 Rattus no
4	17.8	84.8	2583	10	AF302075 Mus muscu
5	17.8	84.8	2601	10	AF157106 Mus muscu
6	17.8	84.8	2652	10	AF302076 Mus muscu
7	17.8	84.8	2694	10	AF302077 Mus muscu
8	17.8	84.8	2892	10	AF157105 Mus muscu
9	17.8	84.8	2925	6	AX033272 Sequence
10	17.8	84.8	2925	10	AF176569 Mus muscu
11	17.8	84.8	83646	8	AB005248 Arabidops
12	17.8	84.8	160601	9	AC021107 Homo sapi
13	17.8	84.8	208249	2	AL607032 Mus muscu
14	17.8	84.8	347950	1	AF003013 Mesorhizo
15	17.4	82.9	62500	2	AC097897 Rattus no
16	17.4	82.9	83610	2	AC097897 Rattus no
17	17.4	81.0	110000	2	AC108277_2 Continuation (3 of
18	17	81.0	159593	9	AC004832 Homo sapi
19	17	81.0	160380	9	AC112777 Homo sapi
20	16.8	80.0	1599	9	AK057750 Homo sapi
21	16.8	80.0	1839	9	AK096095 Homo sapi
22	16.8	80.0	2167	14	AB012663 Homo sapi
23	16.8	80.0	2301	9	AK098799 Homo sapi
24	16.8	80.0	2708	9	AK095645 Homo sapi
25	16.8	80.0	2841	9	AK057670 Homo sapi
26	16.8	80.0	2990	9	AK093663 Homo sapi
27	16.8	80.0	3328	9	HS077948 Human Bruto
28	16.8	80.0	3360	9	AF036613 Homo sapi
29	16.8	80.0	4360	9	AF038969 Homo sapi
30	16.8	80.0	4420	9	AF038968 Homo sapi
31	16.8	80.0	4423	9	AF015553 Homo sapi
32	16.8	80.0	4423	9	AF038967 Homo sapi
33	16.8	80.0	4483	9	AF035737 Homo sapi
34	16.8	80.0	39329	9	AC008995 Homo sapi
35	16.8	80.0	52065	9	AL353576 Human DNA
36	16.8	80.0	56222	9	AL591848 Human DNA
37	16.8	80.0	57245	2	AC068263 Homo sapi
38	16.8	80.0	60442	2	AC115105 Homo sapi
39	16.8	80.0	62500	2	AC103838 Homo sapi
40	16.8	80.0	77745	2	AC087815 Homo sapi
41	16.8	80.0	92858	2	AC024959 Homo sapi
42	16.8	80.0	92858	2	AC024959 Homo sapi
43	16.8	80.0	98485	9	AC107308 Homo sapi
44	16.8	80.0	103872	9	AC026694 Homo sapi
45	16.8	80.0	108490	2	AC091803 Homo sapi

ALIGNMENTS

RESULT 1

AX014709

LOCUS

DEFINITION

AX014709

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AX014709 Sequence 9 from Patent WO953077.

AX014709

AX014709.1 GI:10040982

synthetic construct.

artificial sequences.

1 (bases 1 to 21)

Schwartz,J.C., Gros,C.,

Facchinetti,P.

Novel nep ii membrane metalloprotease and its use for screening

inhibitors useful in therapy

21 bp

DNA

linear

PAT 07-SEP-2000

JOURNAL Patent: WO 9953077-A 9 21-OCT-1999;  
 INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
 CLAUDE (FR); OUIMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
 MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

FEATURES Location/Qualifiers  
 1. 21  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="oligonucleotide"  
 5 a 7 c 5 g 4 t

BASE COUNT 5 a 7 c 5 g 4 t

ORIGIN  
 Query Match 100.0%; Score 21; DB 6; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 1.8;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTACCCCAAGCTGCGTGATAG 21  
 |||||  
 Db 1 CTACCCCAAGCTGCGTGATAG 21

RESULT 2  
 AX014701  
 LOCUS  
 DEFINITION Sequence 1 from Patent WO9953077. PAT 07-SEP-2000  
 AX014701  
 VERSION  
 KEYWORDS  
 SOURCE black rat.  
 ORGANISM Rattus rattus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus  
 1 (bases 1 to 2765)  
 Schwartz,J.C., Gros,C., Oulmet,T., Rose,C., Bonhomme,M.C. and  
 Facchinetti,P.  
 Novel nep ii membrane metalloprotease and its use for screening  
 inhibitors useful in therapy  
 Patent: WO 9953077-A 1 21-OCT-1999;  
 INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
 CLAUDE (FR); OUIMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
 MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

FEATURES Location/Qualifiers  
 1. 2765  
 /organism="Rattus rattus"  
 /db\_xref="taxon:10117"  
 107. 2431  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC07576.1"  
 /db\_xref="GI:10040976"  
 /translation="MGKSESSGMMERADNCGRRRLGFEVCGLLVLLLLMGAIVTL  
 GVFTSIGKPLNLSLHVSRHRTVVRKVLKRDSSQKSDICTTPSCVIAAAILQNM  
 QSKPCDPIYQACGWLRRHVIPETNRSYFVDILRDEVLKGLVLEDSVQHRPA  
 VEKATLYRSCNQSVIEKRDSEPLNLDVIMDGWPVAMDKNWETGPKWLEROLAV  
 LNSQFNRRVLDLFTWDDQNSRHVYIDOPTLGPMSREYFKEDSHRVAYLQFM  
 TSVATMLRDLNLPGETDLVOEMAQVLEHLEFLANATVPQEKRDVYALYHMGLEE  
 LQERGLKCFNTEFLQNVLSVQVELLPNEVVYGIPILENLEEIDVFPQTQLN  
 IYLVRLVDRIGSQSREKARVDYKALYGTMTMEEVRRKCKVAVNMSAVGSLY  
 IKRAFSDKSIYSLEIKIRSFVDNLDELNMDESKKKAQERKALNREIQIGPYDI  
 ILEONRHLDEYSLSLTFSEDLIFENGQNLKNAORSLLKLRKVDQNLWITGAAYV  
 NAFYSPNRLIVFPAGILQPPFSDQPOALNFGGIVGHEITHGPDNDKFN  
 GMLDWSNFSARHPROQSCMIYQSYNFSWELADQNVNFGSTIGENIADNGGVRQA  
 YKALOWLAEGRRQRLPGLNLTAQLFFINAYQVWCGSYRPEFAIQSIKTDVHPSLK  
 YRVLSQNLPGFSEAFHCPRGSPMHPNRCRIW"

BASE COUNT 684 a 735 c 787 g 559 t

ORIGIN  
 Query Match 100.0%; Score 21; DB 6; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 1.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CTACCCCAAGCTGCGTGATAG 21

Db 363 CTACCCCAAGCTGCGTGATAG 383  
 |||||  
 RESULT 3  
 AC094732 174953 bp DNA linear HTG 20-DEC-2001  
 LOCUS Rattus norvegicus clone CH230-516, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION 56 unordered pieces.  
 AC094732  
 VERSION AC094732.2 GI:17941511  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus  
 1 (bases 1 to 174953)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
 Benton,J., Bimager,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
 Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
 Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
 Coyle,M.D., Dathorne,S.R., David,R., David,M.L., Davis,C.,  
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C.,  
 Hollins,L.B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
 Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,  
 Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,  
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
 Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojebokan,I., Rolfe,M.,  
 Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,  
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,  
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,  
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
 Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,  
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 174953)  
 Worley,K.C.  
 Direct Submission  
 Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Dec 20, 2001 this sequence version replaced gi:15624568.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GBGF

Center clone name: CH230-516  
 ----- Summary Statistics  
 findPhrapList  
 Consensus quality: 152255 bases at least Q40  
 Consensus quality: 158448 bases at least Q30  
 Consensus quality: 164461 bases at least Q20  
 Estimated insert size: 155965; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 56 contigs, the true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 13782: contig of 13782 bp in length  
 \* 13783 13882: gap of unknown length  
 \* 13883 23287: contig of 9405 bp in length  
 \* 23288 23387: gap of unknown length  
 \* 23388 28081: contig of 4694 bp in length  
 \* 28082 28181: gap of unknown length  
 \* 28182 33807: contig of 5626 bp in length  
 \* 33808 33907: gap of unknown length  
 \* 33908 39271: contig of 5364 bp in length  
 \* 39272 39371: gap of unknown length  
 \* 39372 44270: contig of 4899 bp in length  
 \* 44271 44370: gap of unknown length  
 \* 44371 47723: contig of 3353 bp in length  
 \* 47724 47823: gap of unknown length  
 \* 47824 53427: contig of 5604 bp in length  
 \* 53428 53527: gap of unknown length  
 \* 53528 57303: contig of 3776 bp in length  
 \* 57304 57403: gap of unknown length  
 \* 57404 62018: contig of 4615 bp in length  
 \* 62019 62118: gap of unknown length  
 \* 62119 66766: contig of 4558 bp in length  
 \* 66767 70201: contig of 3425 bp in length  
 \* 70202 70301: gap of unknown length  
 \* 70302 74082: contig of 3781 bp in length  
 \* 74083 74182: gap of unknown length  
 \* 74183 78306: contig of 4124 bp in length  
 \* 78307 78406: gap of unknown length  
 \* 78407 81222: contig of 2816 bp in length  
 \* 81223 81322: gap of unknown length  
 \* 81323 84552: contig of 3230 bp in length  
 \* 84553 84652: gap of unknown length  
 \* 84653 88839: contig of 4187 bp in length  
 \* 88840 88939: gap of unknown length  
 \* 88940 92256: contig of 3317 bp in length  
 \* 92257 92356: gap of unknown length  
 \* 92357 94995: contig of 2639 bp in length  
 \* 94996 95095: gap of unknown length  
 \* 95096 98465: contig of 3370 bp in length  
 \* 98466 98565: gap of unknown length  
 \* 98566 101532: contig of 3067 bp in length  
 \* 101533 101732: gap of unknown length  
 \* 101733 104939: contig of 3207 bp in length  
 \* 104940 105039: gap of unknown length  
 \* 105040 109240: contig of 4209 bp in length  
 \* 109249 109348: gap of unknown length  
 \* 109349 112435: contig of 3087 bp in length  
 \* 112436 112535: gap of unknown length  
 \* 112536 114676: contig of 2141 bp in length  
 \* 114677 114776: gap of unknown length  
 \* 114777 117567: contig of 2791 bp in length  
 \* 117568 117667: gap of unknown length

\* 117668 119209: contig of 1542 bp in length  
 \* 119210 119309: gap of unknown length  
 \* 119310 121797: contig of 2488 bp in length  
 \* 121798 121897: gap of unknown length  
 \* 121898 124293: contig of 2396 bp in length  
 \* 124294 124393: gap of unknown length  
 \* 124394 126428: contig of 2035 bp in length  
 \* 126429 126528: gap of unknown length  
 \* 126529 129525: contig of 2997 bp in length  
 \* 129526 129625: gap of unknown length  
 \* 129626 132334: contig of 2709 bp in length  
 \* 132335 132434: gap of unknown length  
 \* 132435 135274: contig of 2840 bp in length  
 \* 135275 135374: gap of unknown length  
 \* 135376 138074: contig of 2700 bp in length  
 \* 138075 138174: gap of unknown length  
 \* 138175 139985: contig of 1811 bp in length  
 \* 139986 140085: gap of unknown length  
 \* 140086 142273: contig of 2188 bp in length  
 \* 142274 142374: gap of unknown length  
 \* 142375 143598: contig of 1225 bp in length  
 \* 143599 143698: gap of unknown length  
 \* 143699 145434: contig of 1736 bp in length  
 \* 145435 145534: gap of unknown length  
 \* 145535 146985: contig of 1451 bp in length  
 \* 146986 147085: gap of unknown length  
 \* 147086 148099: contig of 1014 bp in length  
 \* 148100 148199: gap of unknown length  
 \* 148200 150915: contig of 2716 bp in length  
 \* 150916 152015: gap of unknown length  
 \* 152016 152501: contig of 1486 bp in length  
 \* 152502 152601: gap of unknown length  
 \* 152602 154010: contig of 1409 bp in length  
 \* 154011 154110: gap of unknown length  
 \* 154111 155758: contig of 1648 bp in length  
 \* 155759 155858: gap of unknown length  
 \* 155859 157622: contig of 1764 bp in length  
 \* 157623 157722: gap of unknown length  
 \* 157723 159428: contig of 1706 bp in length  
 \* 159429 159528: gap of unknown length  
 \* 159529 161209: contig of 1681 bp in length  
 \* 161210 161309: gap of unknown length  
 \* 161310 163413: contig of 2104 bp in length  
 \* 163414 163513: gap of unknown length  
 \* 163514 164702: contig of 1189 bp in length  
 \* 164703 164802: gap of unknown length  
 \* 164803 165998: contig of 1196 bp in length  
 \* 165999 166098: gap of unknown length  
 \* 166099 167412: contig of 1314 bp in length  
 \* 167413 167512: gap of unknown length  
 \* 167513 169231: contig of 1719 bp in length  
 \* 169232 169331: gap of unknown length  
 \* 169332 170534: contig of 1203 bp in length  
 \* 170535 170634: gap of unknown length  
 \* 170635 172047: contig of 1413 bp in length  
 \* 172048 172147: gap of unknown length  
 \* 172148 173509: contig of 1362 bp in length  
 \* 173510 173609: gap of unknown length  
 \* 173610 174953: contig of 1344 bp in length

## FEATURES

Location/Qualifiers

Query Match 100.0%; Score 21; DB 2; Length 174953;  
 Best Local Similarity 100.0%; Pred. No. 1.6;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGGTAG 21  
 |||||

DB 41361 CTACCCCAAGCTCGGTAG 41341

## RESULT 4

AF302075

LOCUS

2583 bp mRNA linear ROD 11-JUN-2001

```

DEFINITION Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.
ACCESSION AF302075
VERSION AF302075.1 GI:10505359
KEYWORDS
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2583)
AUTHORS Shirotani, K., Tsubuki, S., Iwata, N., Takaki, Y., Harigaya, W., Maruyama, K., Kiryu-Seo, S., Kiyama, H., Iwata, H., Tomita, T., Iwatsubo, T., and Saido, T.C.
TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most rapidly and efficiently among thiorphan- and phosphoramidon-sensitive endopeptidases
J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE 21293028
PUBMED 11278416
REFERENCE 2 (bases 1 to 2583)
AUTHORS Shirotani, K. and Saido, T.C.
TITLE Direct Submission
Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain Science Institute, RIKEN, 2-1 Hirotsawa, Wako-shi, Saitama 351-0198, Japan
FEATURES
source 1. 2583
/organism="Mus musculus"
/db_xref="taxon:10090"
25..2253
CDS
/note="endopeptidase"
/codon_start=1
/product="neprilysin-like peptidase alpha"
/db_xref="GI:10505360"
/translation="MVERAGWCRKSPGFVEYGLMVLILLGAILVTLGVFYSLARD
SLKSDICTPSCVIAAARILENMQSRNCFYQACGWLHRHVIPTNSRYSVF
DILRDEVLKGLVEDTSOHPRAVEKATLYRSCMNQSVIEKRDSEPLLSVLKMGV
GHPVALDKWNETMGLKWELEQLAVNSQFNRRVLDLFWNDQNSRHHVLDQPT
LGMPSREYTFQEDNNHKRYKAYPEFTSVATMLKQDNLSKESAMREMAEVLLEET
HLANATVPOEKRHDTALYHRMDLMELQERGLKGNWTLFIQNLVSSVEVLFPPDEE
VVGYPYLENLEDDISYSARTMNYLVRLVLDIGLSQRFKARVDYRKALYGT
WDEEESKKAQEKAMNIRQIGPYDILEDNKKHDEEYSLTFYEDLVFENGLOLNK
NNAQSLKLRKVDONLALIGAAVNAYSPNNOIVFPAGILOPFFSKDQPOSIN
FGGIGNVGHETHTGDDNGRFDKNGMLDWSNFSARHFOQSOQCMYQYGNFSWE
LADNQNVGFTLGENIADNGVRQAYKAYLRWLADGGKQRLPGLNLTYAQLFFINY
AQVCGSYRPEFAVQSIKTDVHSPKLYRVLGSLQNLPGFSEAFHCPRGSPMPMKRCR
IW"
BASE COUNT 665 a 667 c 736 g 515 t
ORIGIN
Query Match 84.8%; Score 17.8; DB 10; Length 2583;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTACCCCAAGCTCGTGATAG 21
Db 182 CCACCCCAAGCTGTGTGATAG 202
RESULT 5
AF157106
LOCUS AF157106 2601 bp mRNA linear ROD 25-NOV-1999
DEFINITION Mus musculus soluble secreted endopeptidase delta mRNA, alternatively spliced product, complete cds.
ACCESSION AF157106
VERSION AF157106.1 GI:6467400
KEYWORDS
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2601)
AUTHORS Ikeda, K., Emoto, N., Raharjo, S.B., Nurhantari, Y., Saiki, K., Yokoyama, M., and Matsuo, M.
TITLE Molecular identification and characterization of novel membrane-bound metalloprotease, the soluble secreted form of which hydrolyzes a variety of vasoactive peptides
J. Biol. Chem. 274 (45), 32469-32477 (1999)
MEDLINE 20011457
PUBMED 10542292
REFERENCE 2 (bases 1 to 2601)
AUTHORS Ikeda, K., Emoto, N. and Matsuo, M.
TITLE Direct Submission
Submitted (08-JUN-1999) International Center for Medical Research, Kobe University School of Medicine, 7-5-1 Kusunoki, Chuo, Kobe 6500017, Japan
FEATURES
source 1. 2601
/organism="Mus musculus"
/db_xref="taxon:10090"
59..2287
CDS
/note="SEP(delta); metalloprotease; alternatively spliced"
/codon_start=1
/product="soluble secreted endopeptidase delta"
/protein_id="AAF13153.1"
/db_xref="GI:6467401"
/translation="MVERAGWCRKSPGFVEYGLMVLILLGAILVTLGVFYSLARD
SLKSDICTPSCVIAAARILENMQSRNCFYQACGWLHRHVIPTNSRYSVF
DILRDEVLKGLVEDTSOHPRAVEKATLYRSCMNQSVIEKRDSEPLLSVLKMGV
GHPVALDKWNETMGLKWELEQLAVNSQFNRRVLDLFWNDQNSRHHVLDQPT
LGMPSREYTFQEDNNHKRYKAYPEFTSVATMLKQDNLSKESAMREMAEVLLEET
HLANATVPOEKRHDTALYHRMDLMELQERGLKGNWTLFIQNLVSSVEVLFPPDEE
VVGYPYLENLEDDISYSARTMNYLVRLVLDIGLSQRFKARVDYRKALYGT
WDEEESKKAQEKAMNIRQIGPYDILEDNKKHDEEYSLTFYEDLVFENGLOLNK
NNAQSLKLRKVDONLALIGAAVNAYSPNNOIVFPAGILOPFFSKDQPOSIN
FGGIGNVGHETHTGDDNGRFDKNGMLDWSNFSARHFOQSOQCMYQYGNFSWE
LADNQNVGFTLGENIADNGVRQAYKAYLRWLADGGKQRLPGLNLTYAQLFFINY
AQVCGSYRPEFAVQSIKTDVHSPKLYRVLGSLQNLPGFSEAFHCPRGSPMPMKRCR
IW"
BASE COUNT 655 a 681 c 748 g 517 t
ORIGIN
Query Match 84.8%; Score 17.8; DB 10; Length 2601;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTACCCCAAGCTCGTGATAG 21
Db 216 CCACCCCAAGCTGTGTGATAG 236
RESULT 6
AF302076
LOCUS AF302076 2652 bp mRNA linear ROD 11-JUN-2001
DEFINITION Mus musculus neprilysin-like peptidase beta mRNA, complete cds.
ACCESSION AF302076
VERSION AF302076.1 GI:10505361
KEYWORDS
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2652)
AUTHORS Shirotani, K., Tsubuki, S., Iwata, N., Takaki, Y., Harigaya, W., Maruyama, K., Kiryu-Seo, S., Kiyama, H., Iwata, H., Tomita, T., Iwatsubo, T., and Saido, T.C.
TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most rapidly and efficiently among thiorphan- and phosphoramidon-sensitive endopeptidases
J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE 21293028
PUBMED 11278416
REFERENCE 2 (bases 1 to 2652)
AUTHORS Shirotani, K. and Saido, T.C.

```

```

/codon_start=1
/product="neprilysin-like peptidase gamma"
/protein_id="AAG18448.1"
/db_xref="GI:10505364"
/translation="MVERAGWCRKKSPGFVEYGLWVLLILLILLCALVTGLGVFSIALRD
SSLKSDICTPSCVIAARLENNWDSRNPENFYQYACGWLHRHVIPETNSRYSVF
DILRDELIVILKGVLEDTSQHPRAVEKARTIYRSCMNSVIEKRDSEPLLSVLKMG
GWPMADKMNETHGLKWEILERLAVLNSQFNRRVLDLFIWDDQNSSRHVYIDQPT
GLMPSREYFQEDNNHKVRKAYLEFNTSVATMLKDONLSKESAPVREMAQVLELE
HLNATVPQEKRDHDTALYHRMDLMELQRFGLKDRVSLCSPGCTHSDQAGLEGL
NPWADSRLVGLKDFGNWTLFIQNVLSSEVELFPDEEVVYGYPIYLENLELIDSYSA
RTMNSYLVRLVLDIGISLSQRFKARVDYRKALVGTVEEVWRRECYSIVANNMESA
VGSYLKRAFESKDSKSTVRELIEKRSFVDNDELNMDEESKKKAQKAKMNIREQI
GYPIYIDONNHLDDEYSSLIIFYEDLIFENGLOLNKNAQRSLKRLKEKVDQNLWII
GYDVAWFYSPNRRQITPAGLIQPPFSGDQPSQNLNFGGIGWIGHEITHGFDONGR
NFDKNGMLMDWMSNFSAHFQOSQCMIIYOYGNFWELADNONGVNGFSLGNIADNG
QGVQAYKALRWLADGCKDQRLPGLNLTVAQLFFNIYAQVWCGSYRPEPAVQSITKDY
HSPKLYRVLGSQNLPGFSEAFCHPCRGSPHMKRCRIW"
BASE COUNT      686 a 700 c 766 g 542 t
ORIGIN

Query Match      84.8%; Score 17.8; DB 10; Length 2694;
Best Local Similarity 90.5%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CTACCCCAAGCTGCGTGATAG 21
          | | | | | | | | | | | | | | |
Db      182 CCACCCCAAGCTGTGTGATAG 202

RESULT 8
AF157105 LOCUS
DEFINITION Mus musculus soluble secreted endopeptidase mRNA, complete cds.
ACCESSION AF157105
VERSION AF157105.1 GI:6467398
KEYWORDS
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2892)
AUTHORS Ikeda,K., Emoto,N., Raharjo,S.B., Nurhantari,Y., Saiki,K.,
Yokoyama,M. and Matsuo,M.
Molecular identification and characterization of novel
membrane-bound metalloprotease, the soluble secreted form of which
hydrolyzes a variety of vasoactive peptides
J. Biol. Chem. 274 (45), 32469-32477 (1999)
20011457
MEDLINE 10542292
PUBMED 2 (bases 1 to 2892)
REFERENCE Ikeda,K., Emoto,N. and Matsuo,M.
AUTHORS Direct Submission
TITLE Submitted (08-JUN-1999) International Center for Medical Research,
Kobe University School of Medicine, 7-5-1 Kusunoki, Chuoh, Kobe
6500017, Japan
FEATURES.
source Location/Qualifiers
1. 2892
/organism="Mus musculus"
/db_xref="taxon:10090"
281. 2578
/note="SEP; metalloprotease"
/codon_start=1
/product="soluble secreted endopeptidase"
/protein_id="AAF13152.1"
/db_xref="GI:6467399"
/translation="MVERAGWCRKKSPGFVEYGLWVLLILLILLCALVTGLGVFSIGKQL
PLLSLLHFWNDERTVVKRALRDSLSKSDICTPSCVIAARLENNWDSRNPENFY
QYACGWLHRHVIPETNSRYSVFDILRDELIVILKGVLEDTSQHPRAVEKARTIYR
CMNSVIEKRDSEPLLSVLKMGWPNVALKWNETHGLKWEILERLAVLNSQFNRRV
IDLFIWDDQNSSRHVYIDQPTGLMPSREYFQEDNNHKVRKAYLEFNTSVATMLR
DONLSKESAPVREMAQVLELEHLNATVPQEKRDHDTALYHRMDLMELQRFGLK
FNWTLFIQNVLSSEVELFPDEEVVYGYPIYLENLELIDSYRTMNSYLVRLVLD

```







repeat_region	/organism="Homo sapiens"	repeat_region	/rpt_family="Alu"
repeat_region	/db_xref="taxon:9606"	repeat_region	33273..33454
repeat_region	/chromosome="Y"	repeat_region	/rpt_family="MERL_type"
repeat_region	/map="Y"	repeat_region	33815..34113
repeat_region	/clone="RP11-178M5"	repeat_region	/rpt_family="Alu"
repeat_region	/clone_lib="RPC1-11"	repeat_region	34207..34290
repeat_region	35..369	repeat_region	/rpt_family="L2"
repeat_region	/rpt_family="ERV1"	repeat_region	34587..34892
repeat_region	1112..1374	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="ERVK"	repeat_region	35271..35559
repeat_region	1375..1662	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="Alu"	repeat_region	35560..35867
repeat_region	1663..1997	repeat_region	/rpt_family="Alu"
repeat_region	/rpt_family="ERVK"	repeat_region	35868..36340
repeat_region	3761..4402	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="ERVK"	repeat_region	36434..36824
repeat_region	6518..7103	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="ERVK"	repeat_region	36825..37325
repeat_region	9288..9666	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="MER22"	repeat_region	37326..37558
repeat_region	10238..10314	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="MER22"	repeat_region	37604..37674
repeat_region	11805..12194	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="MER22"	repeat_region	38214..38566
repeat_region	14426..14677	repeat_region	/rpt_family="MaLR"
repeat_region	/rpt_family="MER22"	repeat_region	38675..38985
repeat_region	15164..15637	repeat_region	/rpt_family="Alu"
repeat_region	/note="similar to EST A1142349 (NID:g3658708) q963g10.r1"	repeat_region	39021..39325
repeat_region	15309..15734	repeat_region	/rpt_family="Alu"
repeat_region	/note="similar to EST AA936316 (NID:g3094234) on70d04.s1"	repeat_region	39326..39426
repeat_region	16703..17078	repeat_region	/rpt_family="L2"
repeat_region	/rpt_family="MER22"	repeat_region	39807..40092
repeat_region	17572..17629	repeat_region	/note="similar to EST A1805309 (NID:g5391875) te32g02.x1"
repeat_region	/rpt_family="MER22"	repeat_region	40313..40719
repeat_region	17796..17951	repeat_region	/note="similar to EST AA644685 (NID:g2569903) ab87d10.s1"
repeat_region	/rpt_family="MER22"	repeat_region	40494..40794
repeat_region	18443..18831	repeat_region	/note="similar to EST AA971069 (NID:g3146359) op69h07.s1"
repeat_region	/rpt_family="MER22"	repeat_region	42669..43050
repeat_region	19310..19682	repeat_region	/rpt_family="L1"
repeat_region	/rpt_family="MER22"	repeat_region	
repeat_region	19690..20132	repeat_region	
repeat_region	/note="similar to EST AA461577 (NID:g2185441) zx51b02.r1"	repeat_region	
repeat_region	20099..20556	repeat_region	
repeat_region	/note="similar to EST AW181908 (NID:g6450368) xj68c06.x1"	repeat_region	
repeat_region	20289..20702	repeat_region	
repeat_region	/note="similar to EST A1147341 (NID:g3675023) q963g10.s1"	repeat_region	
repeat_region	20333..20763	repeat_region	
repeat_region	/note="similar to EST AW103851 (NID:g6074586) xd76d05.x1"	repeat_region	
repeat_region	21888..22269	repeat_region	
repeat_region	/rpt_family="MER22"	repeat_region	
repeat_region	23759..23835	repeat_region	
repeat_region	/rpt_family="MER22"	repeat_region	
repeat_region	24317..24849	repeat_region	
repeat_region	/rpt_family="MER22"	repeat_region	
repeat_region	24587..24976	repeat_region	
repeat_region	/note="similar to EST AA969211 (NID:g3144391) op51d12.s1"	repeat_region	
repeat_region	26678..27212	repeat_region	
repeat_region	/rpt_family="MER22"	repeat_region	
repeat_region	27958..28038	repeat_region	
repeat_region	/rpt_family="L2"	repeat_region	
repeat_region	28058..28260	repeat_region	
repeat_region	/rpt_family="ERV1"	repeat_region	
repeat_region	28809..28869	repeat_region	
repeat_region	/rpt_family="L2"	repeat_region	
repeat_region	30855..31007	repeat_region	
repeat_region	/rpt_family="L2"	repeat_region	
repeat_region	31236..31374	repeat_region	
repeat_region	/rpt_family="L2"	repeat_region	
repeat_region	31908..31966	repeat_region	
repeat_region	/rpt_family="L2"	repeat_region	
repeat_region	32558..32855	repeat_region	
repeat_region	/rpt_family="Alu"	repeat_region	
repeat_region	32895..33172	repeat_region	

```

----- Project Information
Center project name: BM15L19
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator; 11% of reads
Dye-terminator: Big Dye; 88% of reads
Consensus quality: 207834 bases at least Q40
Consensus quality: 207956 bases at least Q30
Consensus quality: 208060 bases at least Q20
Insert size: 208149; sum-of-contigs
Insert size: 219051; 2.3% error; agarose-fp
Quality coverage: 12.41x in Q20 bases; sum-of-contigs Quality
coverage: 12.08x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 137303: contig of 137303 bp in length
* 137304 137403: gap of 100 bp
* 137404 208249: contig of 70846 bp in length.
FEATURES
    source
        1..208249
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="4"
            /clone_lib="RP23-15L19"
            /clone="RPCI-23"
            /cdate="137303"
            /note="assembly_fragment:04071"
            137404..208249
            /note="assembly_fragment:01838"
            clone_end:T7
            vector_side:right
            54098 a 53647 c 51328 g 49076 t 100 others
BASE COUNT
ORIGIN
    Query Match      84.8%; Score 17.8; DB 2; Length 208249;
    Best Local Similarity 90.5%; Pred. No. 97;
    Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATAG 21
    | ||||| ||||| |||||
Db 170339 CCACCCCAAGCTGTGTGATAG 170359

RESULT 14
AP003013
LOCUS
DEFINITION
    Mesorhizobium loti DNA, complete genome, section 20/21.
ACCESSION
    AP003013.BA000012
VERSION
    AP003013.2 GI:14027324
KEYWORDS
    Mesorhizobium loti (strain:MAFF303099) DNA.
SOURCE
    Mesorhizobium loti
    Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
    Phyllobacteriaceae; Mesorhizobium.
REFERENCE
    1 (sites)
AUTHORS
    Kaneo, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
    Watanabe, A., Iidesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
    Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
    Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,
    Takeuchi, C., Yamada, M. and Tabata, S.
    Complete genome structure of the nitrogen-fixing symbiotic
    Bacterium Mesorhizobium loti
    DNA Res. 7 (6), 331-338 (2000)
JOURNAL
    MEDLINE
    21082930
REFERENCE
    2 (bases 1 to 347950)

```

```

AUTHORS
    Kaneko, T.
TITLE
    Direct Submission
JOURNAL
    Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
    Institute, The First Laboratory for Plant Gene Research; Yana
    1532-3, Kisarazu, Chiba 292-0812, Japan
    (E-mail:kaneko@kazusa.or.jp,
    URL:http://www.kazusa.or.jp/rhizobase/,
    Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
    On May 11, 2001 this sequence version replaced gi:11994988.
COMMENT
    Location/Qualifiers
FEATURES
    source
        1..347950
            /organism="Mesorhizobium loti"
            /strain="MAFF303099"
            /db_xref="taxon:381"
            complement(78..1037)
            /gene="mlr7923"
            complement(78..1037)
            /gene="mlr7923"
            /codon_start=1
            /transl_table=11
            /product="transcriptional regulator"
            /protein_id="BAB53594.1"
            /db_xref="GI:14027325"
            /translation="MRLLSQVNLNLSKIVESAARHGNETFRAGEQFIFITASVSRQVKS
            LEDQLRFKIFQGGNAVSLTPEGETYVSRVEALERIVAAASMEATGQSOEHLKISVL
            PTEAARWLFLPRLPFRORYPDIVMRVSTSYATHEFSTDSFDLEIRYGQGHFNGLPSDL
            LFRDLTPVCSKLPHEVLGDKPVSKVEPDDLHETLLHSDTCTONWQSWLGFAGASF
            VLSEKSIYDFSCMSYEAANNGMFAYANRAIMASDIRAEKIVAFVAFVHHFNSAGWY
            FVSPHKCLAAARKVLLFKQWMAEALTORQDSEIRDLATEAV"
            1150..1986
            /gene="mlr7924"
            1150..1986
            /gene="mlr7924"
            /note="unknown protein"
            /codon_start=1
            /transl_table=11
            /protein_id="BAB53595.1"
            /db_xref="GI:14027326"
            /translation="MNAHTIPELRYAMSRREALIGHETANKVSSFGVAQVLYLHGYPDAIL
            AAIEAALKLSDSHAVHKLDDLTFTIGADRFIPEIKELLDKRLERLSDMMGTGLEP
            YLSIVGSTVTFTMPRDGAVEMHCVDPVFTLPLSLNPLVGGHLEIYCDSDSETRA
            ILESGREIPRNRIRIDHKMNYATLIGQFLGVLRHTAPIQYGERVTLVLNQRSVAKPY
            DDNRMFYLAADNDHDREWNELAEQVWTNQLPAYRRFEAHEPTPAPASETVQSGGARG
            SW"
            2032..2958
            /gene="mlr7925"
            2032..2958
            /gene="mlr7925"
            /note="hypothetical protein"
            /codon_start=1
            /transl_table=11
            /protein_id="BAB53596.1"
            /db_xref="GI:14027327"
            /translation="MVQHRDAPGRFLFRHGSSISALVFAAGAVALTWSTNALVGSLL
            ATHPVSVQFQFAGALVAFALIRFMSREAPSVAPRAAVALAVGIIGLVGTWVLYQ
            IAFASMPVIEANLVAYTWPLVAAAVIAFDNPRPALLGLAALVGFVGVVALVIGSRE
            NNWFGDLVGYFAAGSALCMAFVSWVGRLATSPDRLLLPSSLVGVALLWSAGEG
            VAWPAGMDLALGILYAGPMGLGYTFWSRALKLEGSGKVAVVAYLTPITASTLLLLFSLG
            EQLTTAAGAVLVIGSCIADVLERSEAQDYV"
            2951..3154
            /gene="msr7926"
            2951..3154
            /gene="msr7926"
            /note="unknown protein"
            /codon_start=1
            /transl_table=11
            /protein_id="BAB53597.1"
            /db_xref="GI:14027328"
            /translation="MFDDEARQEAHWLKEFGAEAPLYAMKAERAEKQDFGRC
            ARWRRIELADARTTKSAASKY"
            3307..3759
            /gene="mlr7927"
            3307..3759

```



Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, J., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

# TITLE JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L21835  
Center clone name: 26\_N\_14

\* NOTE: This record contains 78 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 686: contig of 686 bp in length  
787 786: gap of 100 bp  
787 1508: contig of 722 bp in length  
1509 1608: gap of 100 bp  
1609 2286: contig of 678 bp in length  
2287 2386: gap of 100 bp  
2387 3070: contig of 684 bp in length  
3071 3170: gap of 100 bp  
3171 3868: contig of 698 bp in length  
3869 3968: gap of 100 bp  
3969 4674: contig of 706 bp in length  
4675 4774: gap of 100 bp  
4775 5496: contig of 722 bp in length  
5497 5596: gap of 100 bp  
5597 6304: contig of 708 bp in length  
6305 6404: gap of 100 bp  
6405 7122: contig of 718 bp in length  
7123 7222: gap of 100 bp  
7223 7905: contig of 683 bp in length  
7906 8005: gap of 100 bp  
8006 8709: contig of 704 bp in length  
8710 8809: gap of 100 bp  
8810 9490: contig of 681 bp in length  
9491 9590: gap of 100 bp  
9591 10292: contig of 702 bp in length  
10293 10392: gap of 100 bp  
10393 11112: contig of 720 bp in length  
11113 11212: gap of 100 bp  
11213 11917: contig of 705 bp in length  
11918 12017: gap of 100 bp  
12018 12713: contig of 696 bp in length  
12714 12813: gap of 100 bp  
12814 13511: contig of 698 bp in length  
13512 13611: gap of 100 bp  
13612 14299: contig of 688 bp in length  
14300 14399: gap of 100 bp  
14400 15114: contig of 715 bp in length

15115 15214: gap of 100 bp  
15215 15895: contig of 681 bp in length  
15896 15995: gap of 100 bp  
15996 16704: contig of 709 bp in length  
16705 16804: gap of 100 bp  
16805 17523: contig of 719 bp in length  
17524 17623: gap of 100 bp  
17624 18332: contig of 709 bp in length  
18333 18432: gap of 100 bp  
18433 19115: contig of 683 bp in length  
19116 19215: gap of 100 bp  
19216 19936: contig of 721 bp in length  
19937 20036: gap of 100 bp  
20037 20749: contig of 713 bp in length  
20750 20849: gap of 100 bp  
20850 21541: contig of 692 bp in length  
21542 21641: gap of 100 bp  
21642 22347: contig of 706 bp in length  
22348 22447: gap of 100 bp  
22448 23156: contig of 709 bp in length  
23157 23256: gap of 100 bp  
23257 23934: contig of 678 bp in length  
23935 24034: gap of 100 bp  
24035 24745: contig of 711 bp in length  
24746 24845: gap of 100 bp  
24846 25548: contig of 703 bp in length  
25549 25648: gap of 100 bp  
25649 26364: contig of 716 bp in length  
26365 26464: gap of 100 bp  
26465 27176: contig of 712 bp in length  
27177 27276: gap of 100 bp  
27277 27989: contig of 713 bp in length  
27990 28089: gap of 100 bp  
28090 28776: contig of 687 bp in length  
28777 28876: gap of 100 bp  
28777 29577: contig of 701 bp in length  
29578 29677: gap of 100 bp  
29678 30389: contig of 712 bp in length  
30390 30489: gap of 100 bp  
30490 31190: contig of 701 bp in length  
31191 31290: gap of 100 bp  
31291 31980: contig of 690 bp in length  
31981 32080: gap of 100 bp  
32081 32785: contig of 705 bp in length  
32786 32885: gap of 100 bp  
32886 33596: contig of 711 bp in length  
33597 33696: gap of 100 bp  
33697 34415: contig of 719 bp in length  
34416 34515: gap of 100 bp  
34516 35222: contig of 707 bp in length  
35223 35322: gap of 100 bp  
35323 36033: contig of 711 bp in length  
36034 36133: gap of 100 bp  
36134 36817: contig of 684 bp in length  
36818 36917: gap of 100 bp  
36918 37636: contig of 719 bp in length  
37637 37736: gap of 100 bp  
37737 38450: contig of 714 bp in length  
38451 38550: gap of 100 bp  
38551 39259: contig of 709 bp in length  
39260 39359: gap of 100 bp  
39360 40046: contig of 687 bp in length  
40047 40146: gap of 100 bp  
40147 40824: contig of 678 bp in length  
40825 40924: gap of 100 bp  
40925 41578: contig of 654 bp in length  
41579 41678: gap of 100 bp  
41679 42397: contig of 719 bp in length  
42398 42497: gap of 100 bp  
42498 43213: contig of 716 bp in length  
43214 43313: gap of 100 bp  
43314 44032: contig of 719 bp in length  
44033 44132: gap of 100 bp

\* 44133 44815: contig of 683 bp in length  
\* 44816 44915: gap of 100 bp  
\* 44916 45621: contig of 706 bp in length  
\* 45622 45721: gap of 100 bp  
\* 45722 46425: contig of 704 bp in length  
\* 46426 46525: gap of 100 bp  
\* 46526 47237: contig of 712 bp in length  
\* 47238 47337: gap of 100 bp  
\* 47338 48026: contig of 689 bp in length  
\* 48027 48126: gap of 100 bp  
\* 48127 48839: contig of 713 bp in length  
\* 48840 48939: gap of 100 bp  
\* 48940 49657: contig of 718 bp in length  
\* 49658 49757: gap of 100 bp  
\* 49758 50461: contig of 704 bp in length  
\* 50462 50561: gap of 100 bp  
\* 50562 51270: contig of 709 bp in length  
\* 51271 51370: gap of 100 bp  
\* 51371 52030: contig of 660 bp in length  
\* 52031 52130: gap of 100 bp  
\* 52131 52848: contig of 718 bp in length  
\* 52849 52948: gap of 100 bp  
\* 52949 53662: contig of 714 bp in length  
\* 53663 53762: gap of 100 bp  
\* 53763 54470: contig of 708 bp in length  
\* 54471 54570: gap of 100 bp  
\* 54571 55293: contig of 723 bp in length

Query Match 82.9%; Score 17.4; DB 2; Length 62500;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCCCAAGCTGGGTGATAG 21  
||||||| ||| |||||  
Db 26265 TACCCCAAGCTGGGTGATAG 26284

Search completed: July 8, 2003, 03:34:35  
Job time : 234.102 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1064.65 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-9  
Perfect score: 21  
Sequence: 1 ctacccaagctgctgatag 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_man:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	87.6	289	12 BF830865	BF830865 CMI-HT095
2	18.4	87.6	307	12 BF830867	BF830867 CMI-HT095
3	17.8	84.8	600	13 B1989738	B1989738 4044-75 M
4	17.4	82.9	419	17 A2962129	A2962129 2M0230P14
5	17.4	82.9	466	13 B3208146	B3208146 BJ208146
6	17	81.0	519	12 BE775595	BE775595 MY-04-D-0

176	80.0	16.8	80.0	14 N36750
208	80.0	16.8	80.0	12 BF809553
228	80.0	16.8	80.0	12 BF809530
331	80.0	16.8	80.0	9 AA179588
332	80.0	16.8	80.0	13 BG994036
333	80.0	16.8	80.0	13 BG994389
371	80.0	16.8	80.0	14 C03725
376	80.0	16.8	80.0	12 BF768673
377	80.0	16.8	80.0	9 AA323181
378	80.0	16.8	80.0	14 W80887
391	80.0	16.8	80.0	10 BE169628
394	80.0	16.8	80.0	12 BF378044
395	80.0	16.8	80.0	10 BE350500
398	80.0	16.8	80.0	14 H06435
405	80.0	16.8	80.0	10 W388274
407	80.0	16.8	80.0	9 A1624686
424	80.0	16.8	80.0	13 BM664172
426	80.0	16.8	80.0	9 A1613203
426	80.0	16.8	80.0	10 BE169768
429	80.0	16.8	80.0	10 AV682334
429	80.0	16.8	80.0	12 BF805896
437	80.0	16.8	80.0	9 A1567891
437	80.0	16.8	80.0	14 R10313
437	80.0	16.8	80.0	14 R77694
438	80.0	16.8	80.0	14 R59189
439	80.0	16.8	80.0	17 BH746420
442	80.0	16.8	80.0	9 A1801613
449	80.0	16.8	80.0	14 N29917
455	80.0	16.8	80.0	9 A1129478
455	80.0	16.8	80.0	14 H95940
458	80.0	16.8	80.0	10 AW503886
461	80.0	16.8	80.0	9 AA075937
468	80.0	16.8	80.0	10 AW386247
470	80.0	16.8	80.0	12 BE719507
474	80.0	16.8	80.0	14 BQ548733
478	80.0	16.8	80.0	9 AA741448
480	80.0	16.8	80.0	13 BM661988
487	80.0	16.8	80.0	12 BE812758
488	80.0	16.8	80.0	10 BE464803

ALIGNMENTS

RESULT 1	BF830865	289 bp	linear	EST 13-JAN-2001
LOCUS	CMI-HT0950-031000-465-cl1	HT0950	Homo sapiens	CDNA, mRNA sequence.
DEFINITION	BF830865			
ACCESSION	BF830865.1	GI:12178097		
VERSION	EST.			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 289)			
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.			
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7), 3491-3496	(2000)	
MEDLINE	20202663			
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001			

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1&t2=CM1-HT0950-031000-465-cl1&t3=2000-10-03&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 124  
High quality sequence stop: 289.

## FEATURES

Location/Qualifiers  
1..289  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0950"  
/dev\_stage="Adult"

/note="Organ: head,neck; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 87 a 65 c 57 g 80 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 12; Length 289;  
Best Local Similarity 95.0%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGTGATA 20  
|||||  
DB 125 CTACCCCAAGCTGTGTGATA 144

## RESULT 2

BF830867  
LOCUS CM1-HT0950-031000-465-g01 HT0950 Homo sapiens cDNA, mRNA sequence. EST 13-JAN-2001

DEFINITION BF830867  
ACCESSION BF830867  
VERSION BF830867.1 GI:12178100  
KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

AUTHORS  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## JOURNAL

MEDLINE  
COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM1&t2=CM1-HT0950-031000-465-g01&t3=2000-10-03&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 48  
High quality sequence stop: 166.

## FEATURES

Location/Qualifiers  
1..307  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone\_lib="HT0950"  
/dev\_stage="Adult"

/note="Organ: head,neck; Vector: puc18; Site:1: SmaI;  
Site:2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 92 a 64 c 65 g 86 t  
ORIGIN

Query Match 87.6%; Score 18.4; DB 12; Length 307;  
Best Local Similarity 95.0%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGTGATA 20  
|||||  
DB 143 CTACCCCAAGCTGTGTGATA 162

## RESULT 3

BF989738  
LOCUS 4044-75 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence. EST 20-DEC-2001

DEFINITION BF989738  
ACCESSION BF989738.1 GI:17960740  
VERSION BF989738.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 600)  
Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,  
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)

## REFERENCE

AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center.  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

## FEATURES

Source  
Location/Qualifiers  
1..600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.  
Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps  
(mannatis); Cloning technique: CUA Cloning (CloneAmp,  
Life Technologies); Average insert size: 1.8 Kb;  
Insertion site: TACGCCACTGAATTCGTAGT---> Other  
information regarding entire library may be found at  
http://pga.swmed.edu/Data/Libraries/microarray\_cdna\_librar  
ies.htm.

BASE COUNT 146 a 182 c 157 g 115 t  
ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 600;  
Best Local Similarity 90.5%; Pred. No. 3.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGTGATAG 21  
|||||  
DB 98 CCACCCCAAGCTGTGTGATAG 118

```

RESULT 4
LOCUS      AZ962129/c
DEFINITION 2M0230P14R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
ACCESSION  AZ962129
VERSION     AZ962129.1
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus.
REFERENCE   1 (bases 1 to 419)
AUTHORS     Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
            Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
            M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,
            and Wright, D., Weiss, R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0230 row: P column: 14
            Seq primer: CACACAGGAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 419.
FEATURES   source
            1..419
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGC2M0230P14"
            /clone_lib="Mouse 10kb plasmid UUGC2M library"
            /sex="Female"
            /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
            /note="Vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (female) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
            inducible derivative of plasmid RL. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."
BASE COUNT 129 a 72 c 120 g 98 t
ORIGIN
Query Match      82.9%; Score 17.4; DB 17; Length 419;
Best Local Similarity 94.7%; Pred. No. 4.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TACCCCAAGCTGGTGATA 20
        |||||
Db      292 TACCCCAAGCTGGTGAAA 274
        |||||
REFERENCE
AUTHORS     Kamoun, S., Hraber, P. T., Sobral, B. W. S., Nuss, D. and Govers, F.
TITLE       Initial assessment of gene diversity for the oomycete pathogen

```

```

RESULT 5
LOCUS      BJ208146
DEFINITION BJ208146 Y. Ogiwara unpublished cDNA library, Wh Triticum aestivum
            cDNA clone wh/a03 5', mRNA sequence.
ACCESSION  BJ208146
VERSION     BJ208146.1
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
REFERENCE   1 (bases 1 to 466)
AUTHORS     Ogiwara, Y. and Murai, K.
TITLE       Expressed genes in Triticum aestivum
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
            1..466
            /organism="Triticum aestivum"
            /cultivar="Chinese Spring"
            /db_xref="taxon:4565"
            /clone="wh/a03"
            /clone_lib="Y. Ogiwara unpublished cDNA library, Wh"
            /tissue_type="spike at meiosis"
            /dev_stage="Feekes' scale 9"
            /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
            Site_1: EcoRI; Site_2: XhoI; Plants were grown under
            hydroponic conditions at UC Davis, salt stressed for 12
            hours, and for 7 days, then dissected and frozen (Akhunov
            in J. Dvorak Lab). Total RNA was prepared from sheath
            tissue, equal quantities of RNA were pooled from the two
            samples, polyA was purified from the pooled RNA, a cDNA
            library was made, and the cDNA clones were in vivo
            excised to give phagescript phagemids in the TJ Close lab
            at the University of California, Riverside (Akhunov, Chin
            , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang).
            Plasmid DNA preparations and DNA sequencing were
            performed in the OD Anderson lab (all other authors)."
```

```

BASE COUNT 118 a 132 c 114 g 101 t
ORIGIN
Query Match      82.9%; Score 17.4; DB 13; Length 466;
Best Local Similarity 94.7%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TACCCCAAGCTGGTGATA 20
        |||||
Db      430 TACCCCAAGCTGGTGATA 448
        |||||

```

```

RESULT 6
LOCUS      BE775595
DEFINITION MY-04-D-03 PinfestansMY Phytophthora infestans cDNA, mRNA sequence.
ACCESSION  BE775595
VERSION     BE775595.1
KEYWORDS    EST.
SOURCE      potato late blight agent.
ORGANISM    Phytophthora infestans
REFERENCE   1 (bases 1 to 519)
AUTHORS     Kamoun, S., Hraber, P. T., Sobral, B. W. S., Nuss, D. and Govers, F.
TITLE       Initial assessment of gene diversity for the oomycete pathogen

```



profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 71 a 42 c 39 g 56 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 12; Length 208;  
Best Local Similarity 90.0%; Pred. No. 7.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGGTGATA 20  
|||||  
Db 144 CTACCCCAAGCTGTGTAATA 125

RESULT 9  
BF809530/c  
LOCUS 228 bp mRNA linear EST 12-JAN-2001  
DEFINITION QV0-C10199-161100-506-all C10199 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF809530  
VERSION BF809530.1 GI:12138519  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 228)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-C10199-161100-506-all&t3=2000-11-16&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 228.  
Location/Qualifiers  
1. .228

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="C10199"  
/dev\_stage="Adult"  
/note="Organ: colon\_ins; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 76 a 49 c 44 g 59 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 12; Length 228;  
Best Local Similarity 90.0%; Pred. No. 8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGGTGATA 20

Db 164 CTACCCCAAGCTGTGTAATA 145  
|||||  
|||||

RESULT 10  
AA179588/c  
LOCUS 331 bp mRNA linear EST 31-DEC-1996  
DEFINITION ZP49d10.sl Strataene HeLa cell s3 937216 Homo sapiens cDNA clone IMAGE:612787 3', mRNA sequence.  
ACCESSION AA179588  
VERSION AA179588.1 GI:1760974  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 331)  
AUTHORS Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, D., Rohlfing, T., Schellenberg, K., Soares, M.B., Tab, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

TITLE This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
JOURNAL Seq primer: -40M13 fwd. from Amersham  
MEDLINE High quality sequence stop: 286.  
COMMENT Location/Qualifiers  
1. .331

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="GDB:4644626"  
/db\_xref="taxon:9606"  
/clone="IMAGE:612787"  
/clone\_lib="Stratagene HeLa cell s3 937216"  
/sex="female"  
/dev\_stage="HeLa S3 cell line"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR Vector. -5' adaptor sequence: 5' GAATTCGACGAG 3' -3', adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'."

BASE COUNT 86 a 66 c 71 g 106 t  
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 331;  
Best Local Similarity 90.0%; Pred. No. 8.8e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTCGGTGATA 20  
|||||  
Db 121 CTACCCCAAGCTGTGTAATA 102

RESULT 11  
BG994036  
LOCUS 332 bp mRNA linear EST 13-JUN-2001  
DEFINITION PM0-HT0913-120201-008-g11 HT0913 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BG994036  
VERSION BG994036.1 GI:14398106  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 332)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-HT0913-120201-008-g1&t3=2001-02-12&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 3

High quality sequence stop: 332.

Location/Qualifiers

1. 332

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0913"

/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 86 a 57 c 71 g 118 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 332;

Best Local Similarity 90.0%; Pred. No. 8.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTACCCCAAGCTGGTGATA 20

|||||

Db 148 CTACCCCAAGCTGGTGATA 167

RESULT 12

LOCUS BG994389 333 bp mRNA linear EST 13-JUN-2001

DEFINITION PM0-HT0913-130201-007-d12 HT0913 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG994389

VERSION BG994389.1 GI:14398459

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 333)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-HT0913-120201-008-g1&t3=2001-02-12&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 3

High quality sequence stop: 332.

Location/Qualifiers

1. 332

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0913"

/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 86 a 57 c 71 g 118 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 332;

Best Local Similarity 90.0%; Pred. No. 8.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTACCCCAAGCTGGTGATA 20

|||||

Db 148 CTACCCCAAGCTGGTGATA 167

RESULT 12

LOCUS BG994389 333 bp mRNA linear EST 13-JUN-2001

DEFINITION PM0-HT0913-130201-007-d12 HT0913 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG994389

VERSION BG994389.1 GI:14398459

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 333)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-HT0913-120201-007-d12&t3=2001-02-13&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 26

High quality sequence stop: 333.

Location/Qualifiers

1. 333

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0913"

/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 87 a 57 c 71 g 118 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 333;

Best Local Similarity 90.0%; Pred. No. 8.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTACCCCAAGCTGGTGATA 20

|||||

Db 149 CTACCCCAAGCTGGTGATA 168

RESULT 13

LOCUS C03725 371 bp mRNA linear EST 30-JUL-1996

DEFINITION C03725 Human heart cDNA (YNakamura) Homo sapiens cDNA clone 3NH2079, mRNA sequence.

ACCESSION C03725

VERSION C03725.1 GI:1466976

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 371)

Tanaka, T., Ogiwara, A., Uchiyama, I., Takagi, T., Yazaki, Y. and Nakamura, Y.

Construction of a normalized directionally cloned cDNA library from adult heart and analysis of 3040 clones by partial sequencing

Genomics 35 (1), 231-235 (1996)

96299762

CONTACT: Yusuke Nakamura

Institute of Medical Science

University of Tokyo

4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

Tel: 81-3-5449-5372

Fax: 81-3-5449-5433

Email: yusuke@ims.u-tokyo.ac.jp

Location/Qualifiers

1. 371

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="3NH2079"

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM0&t2=PM0-HT0913-130201-007-d12&t3=2001-02-13&t4=1>)

Seq primer: puc 18 forward

High quality sequence start: 26

High quality sequence stop: 333.

Location/Qualifiers

1. 333

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="HT0913"

/dev\_stage="Adult"

/note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 87 a 57 c 71 g 118 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 13; Length 333;

Best Local Similarity 90.0%; Pred. No. 8.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CTACCCCAAGCTGGTGATA 20

|||||

Db 149 CTACCCCAAGCTGGTGATA 168

RESULT 13

LOCUS C03725 371 bp mRNA linear EST 30-JUL-1996

DEFINITION C03725 Human heart cDNA (YNakamura) Homo sapiens cDNA clone 3NH2079, mRNA sequence.

ACCESSION C03725

VERSION C03725.1 GI:1466976

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 371)

Tanaka, T., Ogiwara, A., Uchiyama, I., Takagi, T., Yazaki, Y. and Nakamura, Y.

Construction of a normalized directionally cloned cDNA library from adult heart and analysis of 3040 clones by partial sequencing

Genomics 35 (1), 231-235 (1996)

96299762

CONTACT: Yusuke Nakamura

Institute of Medical Science

University of Tokyo

4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

Tel: 81-3-5449-5372

Fax: 81-3-5449-5433

Email: yusuke@ims.u-tokyo.ac.jp

Location/Qualifiers

1. 371

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="3NH2079"

```
/clone_lib="Human.heart cdna (YNakamura)"
/dev_stage="adult"
/notes="Organ: heart; normalized directionally cloned cdna
from adult heart"
```

BASE COUNT 121 a 89 c 76 g 82 t 3 others

Query Match 80.0%; Score 16.8; DB 14; Length 371;

Best Local Similarity 90.0%; Pred. No. 9.1e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20

|||||

Db 318 CTACCCCAAGCTGCGTGATA 337

RESULT 14

BE768673/c

LOCUS 376 bp mRNA linear EST 20-SEP-2000

DEFINITION QV2-FT0010-090800-300-c04 FT0010 Homo sapiens CDNA, mRNA sequence.

ACCESSION BE768673

VERSION BE768673.1 GI:10222331

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 376)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-QV2-FT0010-090

800-300-c04st3-2000-08-09st4-1)

Seq primer: puc 18 forward

High quality sequence stop: 375.

Location/Qualifiers

1. 376

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="FT0010"

/dev\_stage="Adult"

/note="Organ: prostate\_tumor; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 122 a 80 c 69 g 105 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 12; Length 376;

Best Local Similarity 90.0%; Pred. No. 9.1e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20

|||||

Db 318 CTACCCCAAGCTGCGTGATA 337

|||||

QY 1 CTACCCCAAGCTGCGTGATA 20

|||||

Db 186 CTACCCCAAGCTGCGTGATA 205

|||||

Search completed: July 8, 2003, 09:21:30

Job time : 1068.8 secs

Db 329 CTACCCCAAGCTGCGTGATA 310

|||||

RESULT 15

AA323181

LOCUS 377 bp mRNA linear EST 20-APR-1997

DEFINITION EST25923 Cerebellum II Homo sapiens CDNA 5' end, mRNA sequence.

ACCESSION AA323181

VERSION AA323181.1 GI:1975506

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 377)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult

, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White

, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,

Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald

, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodak, A.,

Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,

Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,

Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,

Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.

, Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,

Kunsch, C., Hung, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,

Wei, Y.F., Wung, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon

, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and

Venter, J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

Other\_ESTs: THCI70366

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 3018699056

Fax: 3018699423

Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse

FEATURES

Source

Location/Qualifiers

1. 377

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):123711"

/db\_xref="taxon:9606"

/clone\_lib="Cerebellum II"

/tissue\_type="cerebellum"

/dev\_stage="adult"

/notes="Organ: brain; Vector: pBluescript SK-; Site\_1:

ECORI; Site\_2: XhoI"

BASE COUNT 96 a 62 c 74 g 141 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 377;

Best Local Similarity 90.0%; Pred. No. 9.1e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACCCCAAGCTGCGTGATA 20

|||||

Db 186 CTACCCCAAGCTGCGTGATA 205

|||||

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 : Search time 231.102 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-10  
Perfect score: 21  
Sequence: 1 cggcaccatggtatcccgag 21

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2034640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pi.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rod.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgtgo\_hum.\*  
40: em\_hgtgo\_mus.\*  
41: em\_hgtgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	21	100.0	21	6	AX014710	Sequence
2	21	100.0	2765	6	AX014701	AX014710 Sequence
3	21	100.0	174953	2	AC094732	AX014701 Sequence
4	19	90.5	19	6	AX014708	AX014708 Sequence
5	17.8	84.8	349498	1	AP003002	AP003002 Mesorhizo
6	16.8	80.0	1679	10	AF212319	AF212319 Mus muscu
7	16.8	80.0	1863	10	MM051167	U51167 Mus musculu
8	16.8	80.0	2494	10	BC024640	BC024640 Mus muscu
9	16.8	80.0	2583	10	AF302075	AF302075 Mus muscu
10	16.8	80.0	2601	10	AF157106	AF157106 Mus muscu
11	16.8	80.0	2652	10	AF302076	AF302076 Mus muscu
12	16.8	80.0	2694	10	AF302077	AF302077 Mus muscu
13	16.8	80.0	2892	10	AF157105	AF157105 Mus muscu
14	16.8	80.0	2925	6	AX033272	AX033272 Sequence
15	16.8	80.0	2925	10	AF176569	AF176569 Mus muscu
16	16.8	80.0	11036	1	AE012477	AE012477 xanthomon
17	16.8	80.0	11321	1	AE011159	AE011159 Methanosa
18	16.8	80.0	42611	2	AC100090	AC100090 Mus muscu
19	16.8	80.0	44381	2	AC097605	AC097605 Rattus no
20	16.8	80.0	100188	2	AC123202	AC123202 Rattus no
21	16.8	80.0	100466	9	AC113931	AC113931 Homo sapi
22	16.8	80.0	108397	2	AC129988	AC129988 Rattus no
23	16.8	80.0	122318	2	AC115489	AC115489 Rattus no
24	16.8	80.0	132934	2	AC120073	AC120073 Rattus no
25	16.8	80.0	142698	9	AC025429	AC025429 Homo sapi
26	16.8	80.0	157377	2	AC022827	AC022827 Homo sapi
27	16.8	80.0	158698	9	AC096632	AC096632 Homo sapi
28	16.8	80.0	161794	9	AC021755	AC021755 Homo sapi
29	16.8	80.0	163034	2	AC099432	AC099432 Rattus no
30	16.8	80.0	172761	9	AC092764	AC092764 Pan trogl
31	16.8	80.0	179156	2	AC111214	AC111214 Rattus no
32	16.8	80.0	179369	2	AC123498	AC123498 Rattus no
33	16.8	80.0	183886	2	AC114518	AC114518 Rattus no
34	16.8	80.0	187871	9	AC087188	AC087188 Homo sapi
35	16.8	80.0	188336	2	AC101736	AC101736 Mus muscu
36	16.8	80.0	196954	9	AC008267	AC008267 Homo sapi
37	16.8	80.0	197725	2	AL844164	AL844164 Mus muscu
38	16.8	80.0	208249	2	AL607032	AL607032 Mus muscu
39	16.8	80.0	210484	2	AC127928	AC127928 Rattus no
40	16.8	80.0	257617	2	AC122861	AC122861 Mus muscu
41	16.8	80.0	294250	1	AP001517	AP001517 Bacillus
42	16.4	78.1	37476	9	HS316612	AL031709 Human DNA
43	16.4	78.1	62135	2	AC100614	AC100614 Mus muscu
44	16.4	78.1	162740	2	AC108955	AC108955 Rattus no
45	16.4	78.1	173200	2	AC121922	AC121922 Mus muscu

# ALIGNMENTS

RESULT 1	AX014710	Sequence 10 from Patent WO953077.	21 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX014710	Sequence 10 from Patent WO953077.	21 bp	DNA	linear	PAT 07-SEP-2000
DEFINITION	AX014710	Sequence 10 from Patent WO953077.	21 bp	DNA	linear	PAT 07-SEP-2000
ACCESSION	AX014710	Sequence 10 from Patent WO953077.	21 bp	DNA	linear	PAT 07-SEP-2000
VERSION	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000
KEYWORDS	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000
SOURCE	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000
ORGANISM	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000
REFERENCE	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000
AUTHORS	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000
TITLE	AX014710.1	GI:10040983	21 bp	DNA	linear	PAT 07-SEP-2000

synthetic construct.  
artificial construct  
artificial sequences.  
1 (bases 1 to 21)  
Schwartz, J.C., Gros, C., Quimet, T., Rose, C., Bonhomme, M.C. and  
Facchinetti, P.  
Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy

JOURNAL Patent: WO 9953077-A 10 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OULMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACHINETTI PATRICIA (FR)  
FEATURES  
source  
Location/Qualifiers  
1. 21  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="olligonucleotide"  
BASE COUNT 4 a 8 c 6 g 3 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CGGCACCATGTGATCCCGAG 21  
Db 1 CGGCACCATGTGATCCCGAG 21  
RESULT 2  
AX014701  
LOCUS  
DEFINITION Sequence 1 from Patent WO9953077.  
ACCESSION AX014701  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
black rat.  
Rattus rattus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2765)  
AUTHORS Schwartz, J.C., Gros, C., Oulmet, T., Rose, C., Bonhomme, M.C. and  
Fachinetti, P.  
TITLE Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
JOURNAL Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OULMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACHINETTI PATRICIA (FR)  
FEATURES  
source  
Location/Qualifiers  
1. 2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107. 2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSESVGMERADNCGRRRLGFEVCGLLVLLLLMGAIVTL  
GVFYSIGQLPLNSLLHVRHRTVVRKLRDSSQSDICTPTPCVIAAARILQNM  
QSKPCDFNYACGWLHVRHVPETNSRYSVFDILRDEVLKGVLEDSVQHRPA  
VEKARTLRSCMNSQVIEKRDSEPLNVDLMIGWPVAMDKWMTGPKWELEROLAV  
LNSQFNRLRLDLFTWDDONSRRHVIIDOPTLGMPSREYVFEKDSHRVREAVLQFM  
TSVATMLRDLNLQEAQVHLHLEHLANATVPQEKRDHDTALYHRLGLEE  
LQERFLGFWNTLFIQNLVSQVELLPNEEVYVIGIPILENLEEIIDVFPATLQN  
YLWRLLDRIGLSQRFKEARVDYRKALYGTMTBEVRWRECVSYVNMSEAVGSGLY  
IKRAFSDKSTVSELIKIRSVFDNLDELNMDESKKKAQKALNIREQIGYPDY  
ILEDNRLHDEYSLSLTSEDLYFENGQNLKNNARSLKLRKRVQDQNLWIGAAVV  
NAFYSNNRLTVFPALQIPFPFFSKDQALNFGGIVGIVGHEITHGDFDNGRDKN  
GNMLDWNSFARHPRQSCQMIYQVSNFELADNQNNGVSTLGENIADNGVROA  
KYNLQWLAEGRDORLPLGLNTLYAQLFFINVAQVWGSYRPEFALQISIKTDVHSPFK  
YRVGLSQNLPGFSEAFHCPGSPMPHNRCRIV"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 21; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 CGGCACCATGTGATCCCGAG 21

Db 467 CGGCACCATGTGATCCCGAG 487  
RESULT 3  
AC094732/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus norvegicus  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 174953)  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbara, J.,  
Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Haves, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,  
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,  
Jacks, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,  
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,  
Louise, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,  
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,  
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,  
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinsan, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
DIRECT SUBMISSION  
Unpublished  
2 (bases 1 to 174953)  
Worley, K.C.  
DIRECT SUBMISSION  
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15624568.  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GBGF

Center clone name: CH230-516  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329First call to  
 findPhrapList  
 Consensus quality: 152255 bases at least Q40  
 Consensus quality: 158448 bases at least Q30  
 Consensus quality: 164461 bases at least Q20  
 Estimated insert size: 155965; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-fp estimation  
 Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 56 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of 'N', but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 13782: contig of 13782 bp in length  
 \* 13783 13882: gap of unknown length  
 \* 13883 23287: contig of 9405 bp in length  
 \* 23288 23387: gap of unknown length  
 \* 23388 28081: contig of 4694 bp in length  
 \* 28082 28181: gap of unknown length  
 \* 28182 33607: contig of 5626 bp in length  
 \* 33608 33907: gap of unknown length  
 \* 33908 39271: contig of 5364 bp in length  
 \* 39272 39371: gap of unknown length  
 \* 39372 44270: contig of 4899 bp in length  
 \* 44271 44370: gap of unknown length  
 \* 44371 47723: contig of 3353 bp in length  
 \* 47724 47823: gap of unknown length  
 \* 47824 53427: contig of 5604 bp in length  
 \* 53428 53527: gap of unknown length  
 \* 53528 57303: contig of 3776 bp in length  
 \* 57304 57403: gap of unknown length  
 \* 57404 62018: contig of 4615 bp in length  
 \* 62019 62118: gap of unknown length  
 \* 62119 66676: contig of 4558 bp in length  
 \* 66677 66776: gap of unknown length  
 \* 66777 70201: contig of 3425 bp in length  
 \* 70202 70301: gap of unknown length  
 \* 70302 74082: contig of 3781 bp in length  
 \* 74083 74182: gap of unknown length  
 \* 74183 78306: contig of 4124 bp in length  
 \* 78307 78406: gap of unknown length  
 \* 78407 81222: contig of 2816 bp in length  
 \* 81223 81322: gap of unknown length  
 \* 81323 84552: contig of 3230 bp in length  
 \* 84553 84652: gap of unknown length  
 \* 84653 88839: contig of 4187 bp in length  
 \* 88840 88939: gap of unknown length  
 \* 88940 92256: contig of 3317 bp in length  
 \* 92257 92356: gap of unknown length  
 \* 92357 94995: contig of 2639 bp in length  
 \* 94996 95096: gap of unknown length  
 \* 95097 98465: contig of 3370 bp in length  
 \* 98466 98565: gap of unknown length  
 \* 98566 101632: contig of 3067 bp in length  
 \* 101633 101732: gap of unknown length  
 \* 101733 104939: contig of 3207 bp in length  
 \* 104940 105039: gap of unknown length  
 \* 105040 109248: contig of 4209 bp in length  
 \* 109249 109348: gap of unknown length  
 \* 109349 112435: contig of 3087 bp in length  
 \* 112436 112535: gap of unknown length  
 \* 112536 114676: contig of 2141 bp in length  
 \* 114677 114776: gap of unknown length  
 \* 114777 117567: contig of 2791 bp in length  
 \* 117568 117667: gap of unknown length

\* 117668 119209: contig of 1542 bp in length  
 \* 119210 119309: gap of unknown length  
 \* 119310 121797: contig of 2488 bp in length  
 \* 121798 121897: gap of unknown length  
 \* 121898 124293: contig of 2396 bp in length  
 \* 124294 124393: gap of unknown length  
 \* 124394 126428: contig of 2035 bp in length  
 \* 126429 126528: gap of unknown length  
 \* 126529 129523: contig of 2997 bp in length  
 \* 129524 129623: gap of unknown length  
 \* 129624 132334: contig of 2709 bp in length  
 \* 132335 132434: gap of unknown length  
 \* 132435 135274: contig of 2840 bp in length  
 \* 135275 135374: gap of unknown length  
 \* 135375 138074: contig of 2700 bp in length  
 \* 138075 138174: gap of unknown length  
 \* 138175 139985: contig of 1811 bp in length  
 \* 139986 140085: gap of unknown length  
 \* 140086 142273: contig of 2188 bp in length  
 \* 142274 142373: gap of unknown length  
 \* 142374 143598: contig of 1225 bp in length  
 \* 143599 143698: gap of unknown length  
 \* 143699 145434: contig of 1736 bp in length  
 \* 145435 145534: gap of unknown length  
 \* 145535 146985: contig of 1451 bp in length  
 \* 146986 147085: gap of unknown length  
 \* 147086 148099: contig of 1014 bp in length  
 \* 148100 148199: gap of unknown length  
 \* 148200 150915: contig of 2716 bp in length  
 \* 150916 151015: gap of unknown length  
 \* 151016 152501: contig of 1486 bp in length  
 \* 152502 152601: gap of unknown length  
 \* 152602 154010: contig of 1409 bp in length  
 \* 154011 154110: gap of unknown length  
 \* 154111 155758: contig of 1648 bp in length  
 \* 155759 155858: gap of unknown length  
 \* 155859 157622: contig of 1764 bp in length  
 \* 157623 157723: gap of unknown length  
 \* 157724 159428: contig of 1706 bp in length  
 \* 159429 159528: gap of unknown length  
 \* 159529 161209: contig of 1681 bp in length  
 \* 161210 161309: gap of unknown length  
 \* 161310 163413: contig of 2104 bp in length  
 \* 163414 163513: gap of unknown length  
 \* 163514 164702: contig of 1189 bp in length  
 \* 164703 164802: gap of unknown length  
 \* 164803 165998: contig of 1196 bp in length  
 \* 165999 166099: gap of unknown length  
 \* 166100 167412: contig of 1314 bp in length  
 \* 167413 167512: gap of unknown length  
 \* 167513 169231: contig of 1719 bp in length  
 \* 169232 169331: gap of unknown length  
 \* 169332 170534: contig of 1203 bp in length  
 \* 170535 170634: gap of unknown length  
 \* 170635 172047: contig of 1413 bp in length  
 \* 172048 172147: gap of unknown length  
 \* 172148 173509: contig of 1362 bp in length  
 \* 173510 173609: gap of unknown length  
 \* 173610 174953: contig of 1344 bp in length.

## FEATURES

## Location/Qualifiers

Query Match 100.0%; Score 21; DB 2; Length 174953;  
 Best Local Similarity 100.0%; Pred. No. 2.7;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGGCACCATGTGATCCCCGAG 21  
 ||||||||||||||||  
 DB 40481 CGGCACCATGTGATCCCCGAG 40461

RESULT 4  
 AX014708/c 19 bp DNA linear PAT 07-SEP-2000  
 LOCUS AX014708

DEFINITION Sequence 8 from Patent WO953077.

AX014708  
AX014708.1 GI:10040981

synthetic construct.  
synthetic construct  
artificial sequences.

ORGANISM

REFERENCE

AUTHORS Schwartz,J.C., Gros,C., Oulmet,T., Rose,C., Bonhomme,M.C. and

Fachinetti,P.

TITLE Novel nep-11 membrane metalloprotease and its use for screening

inhibitors useful in therapy

JOURNAL Patent: WO 953077-A 8 21-OCT-1999;

INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS

CLAUDE (FR); OULMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME

MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

Location/Qualifiers

FEATURES

source

1..19 /organism="synthetic construct"

/db\_xref="taxon:32630"

/note="oligonucleotide"

3 a 5 c 8 g 3 t

BASE COUNT

ORIGIN

Query Match 90.5%; Score 19; DB 6; Length 19;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCG 19

|||||

Db 19 CGGCACCATGTGATCCCG 1

RESULT 5

AP003002/c

LOCUS AP003002 349498 bp DNA linear BCT 15-MAY-2001

DEFINITION Mesorhizobium loti DNA, complete genome, section 9/21.

ACCESSION AP003002 BAA000012

VERSION AP003002.2 GI:14023709

KEYWORDS

SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.

ORGANISM

Mesorhizobium loti

Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

Phyllobacteriaceae; Mesorhizobium.

REFERENCE

AUTHORS

Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,

Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,

Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,

Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,

Takeuchi,C., Yamada,M. and Tabata,S.

TITLE Complete genome structure of the nitrogen-fixing symbiotic

bacterium Mesorhizobium loti

JOURNAL bacterium Mesorhizobium loti

MEDLINE DNA Res. 7 (6), 331-338 (2000)

REFERENCE 21082930

2 (bases 1 to 349498)

AUTHORS Kaneko,T.

Direct Submission

TITLE Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research

Institute, The First Laboratory for Plant Gene Research; Yana

1532-3, Kisarazu, Chiba 292-0812, Japan

(E-mail:kaneko@kazusa.or.jp,

URL:http://www.kazusa.or.jp/rhizobase/

Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)

On May 11, 2001 this sequence version replaced gi:11994977.

COMMENT

FEATURES

Source

1..349498

/organism="Mesorhizobium loti"

/strain="MAFF303099"

/db\_xref="taxon:381"

complement(26..1600)

gene

CDS

complement(26..1600)

/gene="mll3424"

/gene="mll3424"

/codon\_start=1  
/transl\_table=1  
/product="transmembrane efflux protein"  
/protein\_id="BAB50315.1"  
/db\_xref="GI:14023710"  
/translation="MHSNFAAIIPLTSRLTNGTFCPOSRRFVLIILASALGFID  
GSILATAMPARIVNLGSLAEAOISNAYALTLSALILAGGAAGDRFLRRFVAGIA  
LFIVASACALAPNAVVLIGFALQIGAAIMVPGSLAIIAKAYPKRGRGAIIGWAA  
ASALTALGPVLGVLVSFGDGIWRAIFAVNPLGLISILYLLVKIPADATKRSLL  
DLGGALATLAFALAYLGLTSMSSSEGHMAGPSIAGAVLLVFLFEQQRQPMID  
LSLRFAGAGANVATFFLYFALSANFLYPLMLIAGWGLSTAEVGFIFLPLSLIAL  
LSPGAGQSDRIGRFRFIAGSLIVAFAGLALLSHAGIHFFHTGILPLMALGIMG  
ALVVSPLSTIMTAVEDKDTGAASGINNAVSRIGGLIIVAAAGSLATVWYATMLDTSV  
RPGIPGPEPVSTGLPPELEATRLAASDMAFSALSTTAVLCLLAAAILIATVVSQAL  
PWPRIDQSWTDSRS"  
complement(1611..1766)  
/gene="msl3425"  
complement(1611..1766)  
/gene="msl3425"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB50316.1"  
/db\_xref="GI:14023711"  
/translation="MDVSLIFPRLPFGSGKSPMTGICIGTGRILANMLVSRATSCWRI  
VAGALGV"  
1734..2386  
/gene="mlr3426"  
1734..2386  
/gene="mlr3426"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB50317.1"  
/db\_xref="GI:14023712"  
/translation="MKRPLASAGRGRIIVGTGVAIALTLIVLEIRVPAHETVHS  
EHLIVAIAPALAPFLYALSELFTLTFWGOQAQHGIIAKSDRLATLNLCLFLAFIA  
LLFSTDLADFLFRFAYLVNLLMLGVTLPASWRYADKNGFVADDAETKRTV  
YLRIKAQILWAIGALLCITPLLSVGFILVVQLIYATALRGSLLRKVG"  
complement(2360..3877)  
/gene="mll3427"  
complement(2360..3877)  
/gene="mll3427"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB50318.1"  
/db\_xref="GI:14023713"  
/translation="MFTIFALSALATGLYSFGALGAPTODNAPRVVTVTANDGGILVA  
QGNLDIYDARGNRIVDADTGKVIATQPTRLDRALRRELRRELGRAPADDDR  
YLLDDPDQMARFRKQLEEGRVIPPADEYDPNGDMSVEAYPPAPQDDGSDYNSIPD  
APQAPKGTIKRQPLNEAIDPQDLEQVNPDTQASLPDPTGKAAVDPSSLGVRQ  
DVAALQVLLDRGGASPGVIGDFGNSVDKALAYNQITGSLNKSTDTVGIQAAQSG  
GDAFASVYITAEADAAGPYVASIPEDYGOKALCMCYISVTEALAEFRHMDENLKSI  
NKGIDFNRCPTIILKAVNGKLIVSTPVARIADTKTKEVYAYDASGKLVAAVPAFIGSA  
DTSPGTGHAVERSIALDPNTYINPNINFGKQNDKILTIIPGPNPGVSGSVIALDKPT  
YGIHGTPEPSKIGTKRGTLNWDARELAKLVSPGVTVFVFGPSADDFAOQ"  
complement(4026..4388)  
/gene="mll3428"  
complement(4026..4388)  
/gene="mll3428"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAB50319.1"  
/db\_xref="GI:14023714"  
/translation="MTLDDYNGFCASLPAANHVVOWGGAHWKVGTVKFAIGGHDQGO  
QFFVTFKSDIADVLREQQLRPAPYIASGMKWIORTTSQSMDDAALKDYLRESHR  
LVALKTKQARKELGLAS"  
complement(4449..7055)  
/gene="mll3429"  
complement(4449..7055)  
/gene="mll3429"  
/gene="mll3429"

/product="phosphoenolpyruvate-protein phosphotransferase: Ptsp"	
Query Match	84.8%; Score 17.8; DB 1; Length 349498;
Best Local Similarity	90.5%; Pred. No. 1.6e+02;
Matches	19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 CGGCACCATGTGATCCCGAG 21 
Db	299423 CGGCACCATGTGATCCCGAG 299403
RESULT 6	
AF212319/c	1679 bp mRNA linear ROD 07-MAY-2001
LOCUS	Mus musculus NADP+-specific isocitrate dehydrogenase mRNA, complete cds; nuclear gene for mitochondrial product.
DEFINITION	
ACCESSION	AF212319
VERSION	AF212319.1 GI:120033361
KEYWORDS	
SOURCE	Mus musculus.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	
AUTHORS	Jo, S.H., Son, M.K., Koh, H.J., Lee, S.M., Song, I.H., Kim, Y.O., Lee, Y.S., Jeong, K.S., Kim, W.B., Park, J.W., Song, B.J. and Huhe, T.L.
TITLE	Control of mitochondrial redox balance and cellular defense against oxidative damage by mitochondrial NADP+-dependent isocitrate dehydrogenase
JOURNAL	J. Biol. Chem. 275 (19), 16168-16176 (2001)
MEDLINE	21238291
PUBMED	11278619
REFERENCE	2 (bases 1 to 1679)
AUTHORS	Huh, T.L., Jo, S.H. and Son, M.K.
TITLE	Direct Submission
JOURNAL	Submitted (06-DEC-1999) Genetic Engineering, Kyungpook National University, San-kyuk-dong 1370, Taegu 702-701, South Korea
FEATURES	Location/Qualifiers 1..1679 /organism="Mus musculus" /strain="3T3" /db_xref="taxon:10090" 46..1404 /note="IDPM" /codon_start=1 /product="NADP+-specific isocitrate dehydrogenase" /protein_id="AAG43538.1" /db_xref="GI:12003362" /translation="MAGLYRVSLSICRSGSARTWAPALTVPSPEQPRRIYAEKRI ATYKPVSVKMGDEMTRINQWIKELILPHVDVQLKFDGLPNRDQINDQVITDIAL KVCQISVAVKQDQITLPDARVEEFKKMKWSPNGTIRNLGGTVFERPIICKNIPRL VPGWTKPITIGRHAGDOVKATDFVVDVDRAGTKLVFTPKDGSASKEWEYVNPAGVG MGMYNTDESIGFASHCQFSQIKWPLVLTNTILKAYDGRFIDQIEFDKHYKT DEDRKNITWEKRLIDDMVAQLKSGWYACKNDYGDQSDILAQGGFSGLGMLTSVL VCPDQKITEAAHAATVTRHYRHEQKGRPTSTNPASIFATWRTGLEHGRKGLDQDLI RFAPQITLEKVCQTVESGAMTKDLACIHGLSNVKNLNEHFLNTDPLDTIKSNLDRALG KQ"
BASE COUNT	404 C 477 G 360 T

```

ORIGIN
Query Match      80.0%; Score 16.8; DB 10; Length 1679;
Best Local Similarity 90.0%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21
    ||||| ||||| ||||| |||||
Db 1640 GGCACCATGTGATCCCGAG 1621

RESULT 7
NMU51167/c
LOCUS Mus musculus isocitrate dehydrogenase mRNA, complete cds.
DEFINITION Mus musculus isocitrate dehydrogenase mRNA, complete cds.
ACCESSION U51167.
VERSION U51167.1
KEYWORDS GI:1236983
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1863)
Yang, L., Luo, H., Vinay, P. and Wu, J.
Molecular cloning of the cDNA of mouse mitochondrial NADP-dependent
isocitrate dehydrogenase and the expression of the gene during
lymphocyte activation
J. Cell. Biochem. 60 (3), 400-410 (1996)
JOURNAL MEDLINE 97021455
PUBMED 8867815
REFERENCE 2 (bases 1 to 1863)
AUTHORS Luo, H.
Direct Submission
TITLE Submitted (11-MAR-1996) Hongyu Luo, Notre-Dame Research Center,
University of Montreal, 1560 Sherbrooke East, Montreal, Quebec, H2L
4M1, Canada
FEATURES
source Location/Qualifiers
1..1863
/organism="Mus musculus"
/strain="Balb/c"
/db_xref="taxon:10090"
/clone="MIDH 5.1"
/tissue_type="heart"
/dev_stage="adult"
24..1595
/codon_start=1
/product="isocitrate dehydrogenase"
/protein_id="AAC52473.1"
/db_xref="GI:1236984"
/translation="MQLVYMLQIQLSGRGCGRAWPGHEHLSWRRGVDLGDRRGL
LSRFLSPAAVAEAAVEAAASDLACSEWPAAGCELLCRASGSARTWAPALATVPS
WPEQPRRYAEKRIKVEKPVEMDGMTRIIQWFKELKILPHVDVQLKYFDLGLPN
RDOTNDQVITDIALAAQKYSMAVKCATITPDEARVEEFLKKMKWSPNGTIRNLGPT
VEREPTKNIPLVPKWTPTIGRHAGDOYKATDFVVDRACTPKLFTPKNGSSA
KWEVYVPPGGVGMVNTDESIGFAHSCFQYSIQKKNPLYLSTKNILKAYDGRF
KDIFEIFDKHTKTDNRNKIWEHRLIDDMVAQLKSGGFVWACKNDGDVQSDIL
ASRFGSLGMLTVLCPDGKTIKAEAAHGTVTVRHYREHQKQPTSTKGIASIFAMTRG
LEHRGKLDQNDQLIRFAQOTREKVCQTVVEGAMTKDLACGIHGLSNVKNLNEHFLNTTDF
LDTIKSLDRLGKO"
BASE COUNT 436 a 482 c 546 g 399 t
ORIGIN
Query Match      80.0%; Score 16.8; DB 10; Length 1863;
Best Local Similarity 90.0%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21
    ||||| ||||| ||||| |||||
Db 1827 GGCACCATGTGATCCCGAG 1808

RESULT 8
BC024640/c
LOCUS Mus musculus, clone MGC:28430 IMAGE:4038046, mRNA, complete cds.
DEFINITION Mus musculus, clone MGC:28430 IMAGE:4038046, mRNA, complete cds.
ACCESSION BC024640
VERSION BC024640.1
KEYWORDS GI:22137612
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2494)
Straussberg, R.
Direct Submission
TITLE Submitted (01-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 37-Row: b Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
FEATURES
source Location/Qualifiers
1..2494
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="MGC:28430 IMAGE:4038046"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI-CGAP Mam5"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
531..1043
/codon_start=1
/product="Unknown (protein for MGC:28430)"
/protein_id="AAH24640.1"
/db_xref="GI:22137613"
/translation="MSQASRRKHVQVVLGHEWVPSDHOQIVKVLRTPGNNLHEVETA
QGQFLVSPSKYKKNIKRIGFLVDPLEEGEKVKAETSVLCKHVSRLQEGHW
PEAFSEVAEKQNNNRSPQLFPAEPOLSGEGSSDDSLFVNTNRQTHSESESE
EDEEEEEEA"
BASE COUNT 586 a 596 c 633 g 679 t
ORIGIN
Query Match      80.0%; Score 16.8; DB 10; Length 2494;
Best Local Similarity 90.0%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21
    ||||| ||||| ||||| |||||
Db 592 GGCACCATGTGATCCCGAG 573

RESULT 9
AF302075
LOCUS Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.
DEFINITION Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.
ACCESSION AF302075
VERSION AF302075
KEYWORDS GI:2583
SOURCE AF302075
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2583)

```

[illegible]

## JOURNAL

Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain  
Science Institute, RIKEN, 2-1 Hirose, Wako-shi, Saitama  
351-0198, Japan

## FEATURES

## source

Location/Qualifiers

1. 2652

/organism="Mus musculus"

/db\_xref="taxon:10090"

25. .2322

/note="endopeptidase"

/codon\_start=1

/product="neprilysin-like peptidase beta"

/protein\_id="AAG18447.1"

/db\_xref="GI:10505362"

/translation="MYERAGWCKKSPGVEYGLMVLILLGAIIVTLGVFYSIGKQL  
PLTSLHFSWDETRVVKRALDSSLSKDICTPSCVIAAARILENDDSRNCPENFY  
QYACGGWLRHHVLPETNSRYSVDFILKGVLEDSQHPRAVEKATLYRS  
CMNQSIVKRRSEPLLSVLKMGWGPVAMKWNETMGLKWELEQLAVLNSQFNRLV  
IDLFWNDQNSRRHVIYIDQPLGMPREYFOEDNNHVKRAYLEFMTSVATMLRK  
DNLKSAMVREMAEVELETHLANATVPQEKRDHVTALYHRMDLMELQERGLKG  
FNWTLFQNVLSVEVLEFPDEEVVYIGIPYLENLEDDISYSARTMQLVWRLVLD  
RIGLSQREKARVDYRKALYGTVEVRECVSVNSNMESAVGSLYIKRAFSSKDS  
KTVRLIEKRSVFDNLDELNMDEESKKAQEKAMNIREQIGYDYLDELNNKHL  
DEEYSSLYEDLYFENGQNLKNAQSLKREKVDQNLWIGAAVNAFVSPNRN  
QIVFPAGILQPPFFSDQPSLNFSGIGWIGHEITHGEDDNGRNDKNGNMLDWSN  
FSARHFQOQSCMIYQGNFSEWELADNQNNGVSTIGENADNGGVRQAYKAYLRWLA  
DGKQDQRLFLNLTYAQLFFINVAQWCGSYRPEFAVQSIKTDVHSPKRVLGSLON  
LPGFSEAFHCPGSPMPMKRCRIW"

## BASE COUNT

682 a 685 c 755 g 530 t

## ORIGIN

Query Match

Best Local Similarity 80.04; Score 16.8; DB 10; Length 2652;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

2 GGCACCATGATGCCCGAG 21

Db

356 GGCACCATGATGCCCGAG 375

## RESULT 12

## AF302077

## LOCUS

Mus musculus neprilysin-like peptidase gamma mRNA, complete cds.

## DEFINITION

## AF302077

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

## CDS

LOCUS AF176569 2925 bp mRNA linear ROD 11-MAY-2000

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 131.941 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-10

Perfect score: 21

Sequence: 1 cggcaccatgtatcccccag 21

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	20	AAZ28816
2	21	100.0	2765	20	AAZ28810
3	19	90.5	19	20	AAZ28814
4	16.8	80.0	1667	24	ABA95078
5	16.8	80.0	2925	21	AAA63763
6	16.8	80.0	3939	23	AA588996
7	16.8	80.0	3939	23	AA588996
8	16.4	78.1	410	22	AAI90923
9	16.4	78.1	4829	22	AAK84366

10	16.2	77.1	97	22	ABA70793
11	16.2	77.1	97	22	AAK13046
12	16.2	77.1	97	22	AAK44999
13	16.2	77.1	97	22	AAI50967
14	16.2	77.1	97	24	ABSI9244
15	16.2	77.1	460	22	ABA58192
16	16.2	77.1	460	22	AAK06272
17	16.2	77.1	460	22	AAK31934
18	16.2	77.1	460	22	AAI37800
19	16.2	77.1	460	22	ABS06696
20	16.2	77.1	840	23	AA575689
21	16.2	77.1	1150	22	AAI21381
22	16.2	77.1	1402	22	AAI21398
23	16.2	77.1	2076	22	AA597186
24	16.2	77.1	2232	24	AA597186
25	16.2	77.1	2262	22	AA597186
26	16.2	77.1	2318	22	AA597186
27	16.2	77.1	2340	22	AA597186
28	16.2	77.1	2580	24	ABN84280
29	16.2	77.1	2636	22	AA596660
30	16.2	77.1	2663	22	AA596661
31	16.2	77.1	2676	21	AA63764
32	16.2	77.1	2714	22	AA596599
33	16.2	77.1	2893	24	ABN84279
34	16.2	77.1	2893	24	ABN84279
35	16.2	77.1	2953	24	ABK48251
36	16.2	77.1	2975	24	AAI28547
37	15.8	75.2	504	22	AA61121
38	15.8	75.2	1107	22	AA61120
39	15.8	75.2	1107	24	AAI47264
40	15.8	75.2	1107	24	AAI47264
41	15.8	75.2	1324	23	ABLI4849
42	15.8	75.2	1565	23	ABLI2653
43	15.8	75.2	3474	23	ABLI4848
44	15.8	75.2	4021	23	ABLI2652
45	15.4	73.3	592	23	ABLI2855

ALIGNMENTS

RESULT 1  
AAZ28816  
ID AAZ28816 standard; DNA; 21 BP.

AC AAZ28816;  
XX

DT 01-FEB-2000 (first entry)

DE Rat membrane metalloprotease NEPII gene probe #6.

Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
cardiovascular disease; neurodegenerative disease; growth disorder;  
hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.

OS Synthetic.  
OS Rattus rattus.

PN FR2777291-A1.

PD 15-OCT-1999.

PF 08-APR-1998; 98FR-0004389.

PR 08-APR-1998; 98FR-0004389.

PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

PI Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;

PI Schwartz JC;

DR WPI; 1999-593429/51.

disorders related to peptidergic signalling in which NEPII in involved, e.g. cardiovascular or neurodegenerative diseases; growth disorders of endocrine origin; disturbances of the hypothalamic-hypophyseal axis or endocrine disorders.

Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;

Query Match 100.0%; Score 21; DB 20; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 0.86;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCGAG 21  
|||||  
DB 467 CGGCACCATGTGATCCCGAG 487

RESULT 3  
AAZ28814/c  
ID AAZ28814 standard; DNA; 19 BP.  
XX  
XX AAZ28814;  
XX AC AC  
XX  
XX  
XX 01-FEB-2000 (first entry)  
XX DE Rat membrane metalloprotease NEPII gene probe #4.  
XX  
XX Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
XX  
XX Synthetic.  
OS Rattus rattus.  
XX  
XX FR2777291-A1.  
PN  
XX  
XX 15-OCT-1999.  
PD  
XX  
XX 08-APR-1998; 98FR-0004389.  
XX  
XX 08-APR-1998; 98FR-0004389.  
PR  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA  
XX  
XX Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
PI  
XX  
XX WPI; 1999-593429/51.  
DR  
XX  
XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease  
XX  
XX Claim 3; Page 20; 29pp; French.  
PS  
XX  
XX Sequences AAZ28811-Z28827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysine II (NEPII) gene  
CC (AAZ28810). NEPII is involved in (in)activation of neuronal and hormonal  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic-hypophyseal axis or endocrine  
CC disorders.  
XX  
XX Sequence 19 BP; 3 A; 5 C; 8 G; 3 T; 0 other;  
SQ

Query Match 90.5%; Score 19; DB 20; Length 19;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCGC 19

Db 19 CGCACCATGTGATCCCCG 1  
|||||

RESULT 4  
ABA95078/c  
ID ABA95078 standard; cDNA: 1667 BP.

XX ABA95078;  
XX 20-MAY-2002 (first entry);  
XX Mouse IDPm encoding cDNA.

XX Isocitrate dehydrogenase; NADPH; AOS; GSH; IDH; IDPm; IDP; mouse;  
KW mitochondrial NADP+ dependent isocitrate dehydrogenase; vasotropic;  
KW cytoplasmic NADP+ dependent isocitrate dehydrogenase; cerebroprotective;  
KW antioxidant; nicotinamide adenine dinucleotide phosphate; enzyme;  
KW antioxidant system; glutathione; aging; gene; ss.

XX Mus sp.  
XX Key Location/Qualifiers  
FH 46..1404  
FT CDS /\*tag= a  
FT /\*product= "IDPm"

XX WO200208405-A1.  
XX 31-JAN-2002.  
XX 20-JUL-2001; 2001WO-KR01242.  
XX 20-JUL-2000; 2000KR-0041516.

XX (TCBI-) TG BIOTECH INC.  
XX (HUHT/) HUH T.  
XX Huh T, Park J, Lee S, Jo S, Son M;  
XX WPI; 2002-227047/28.  
XX P-PSDB; ABB07641.

XX Isocitrate dehydrogenases and their genes in antioxidant system  
PT suppressive of reactive oxygen species-mediated diseases e.g. ischemic  
PT stroke, catalyses production of nicotinamide adenine dinucleotide  
PT phosphate

XX Claim 10; Page 104-107; 112pp; English.

XX The invention relates to isocitrate dehydrogenases selected from a  
CC cytoplasmic NADP+ dependent isocitrate dehydrogenase (IDPc) or a  
CC mitochondrial NADP+ dependent isocitrate dehydrogenase (IDPm). The  
CC isocitrate dehydrogenases are useful catalysing the production of  
CC nicotinamide adenine dinucleotide phosphate (NADPH), useful in improving  
CC antioxidant activity in cells and tissues. An antioxidant system  
CC (AOS) for increasing cellular reduced glutathione (GSH) level, comprising  
CC IDP or its gene, isocitrate as an enzyme substrate, and NADPH as a  
CC reaction product is also provided. The AOS is useful as a cure for or a  
CC preventive agent for reactive oxygen species-mediated diseases which  
CC include ischemic stroke, aging, and acute pulmonary damage caused by a  
CC provision of high pressure oxygen; for use in suppression of radiation-  
CC caused diseases, which include UV- or X-ray-induced skin damage and  
CC concomitant aging. Materials having decreased IDPm and IDPc gene  
CC expression is useful for screening antioxidants inhibitory of reactive  
CC oxygen species-mediated DNA damage. The present sequence represents a  
CC cDNA encoding the mouse IDPm protein.

XX Sequence 1667 BP; 392 A; 438 C; 477 G; 360 T; 0 other;

Query Match 80.0%; Score 16.8; DB 24; Length 1667;

Best Local Similarity 90.0%; Pred. No. 98;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCCG 21  
Db 1640 GGCACCATGTGATCCCCG 1621  
|||||

RESULT 5  
AAA63763  
ID AAA63763 standard; cDNA: 2925 BP.

XX AAA63763;  
XX 04-DEC-2000 (first entry)  
XX cDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.

XX Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
KW NEP-like enzyme; protein production; protein secretion;  
KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
KW fertility; bone disease; abnormal phosphate metabolism; ss.

XX Mus sp.  
XX Key Location/Qualifiers  
FH 332..2629  
FT CDS /\*tag= a  
FT /\*product= "neutral endopeptidase metalloproteinase-like  
enzyme NL-1"

XX WO200047750-A2.  
XX 17-AUG-2000.  
XX 11-FEB-2000; 2000WO-CA00147.  
XX 11-FEB-1999; 99CA-2260376.

XX (UYMO-) UNIV MONTREAL.  
XX Desgroseillers L, Boileau G;  
XX WPI; 2000-549148/50.  
XX P-PSDB; AAB08130.

XX Novel neutral endopeptidase-like metalloproteinase polypeptides and  
PT polynucleotides, used to screen for related sequences and enzyme  
PT inhibitors, used for the treatment of NL-3 related bone disorders -

XX Disclosure; Fig 3; 59pp; English.

XX The present sequence encodes a murine neutral endopeptidase  
CC metalloproteinase-like enzyme, designated NL-1. The specification  
CC also describes NL-2 and NL-3. The NL enzymes are used to test for  
CC specific inhibitors. The N-terminal region of the enzymes can be used  
CC to promote production and secretion of foreign proteins and active  
CC biopeptides, using chimeric constructs containing the foreign protein  
CC downstream from and in phase with the N-terminal region. The NL enzymes  
CC are have been localised to the brain, and may be useful in the  
CC treatment of neurological diseases such as Alzheimer's disease, pain,  
CC and psychiatric disorders. NL enzymes have also been localised to the  
CC testis and ovaries, and may be used to control fertility. They have  
CC also been localised to bones, and may be used to treat bone diseases,  
CC and abnormal phosphate metabolisms related to improper peptide  
CC processing by the NL-3 enzyme.

XX Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;

Query Match 80.0%; Score 16.8; DB 21; Length 2925;

Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCCG 21  
|||||

663:GGCACCAGTGATCCAGAG 682

Db

## RESULT 6

AAS88996  
ID AAS88996 standard; cDNA; 3939 BP.

XX AC AAS88996;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #24800.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG24809.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity

XX PS Claim 1; SEQ ID No 24800; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human  
XX CC diagnostic coding sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3939 BP; 933 A; 1072 C; 1040 G; 894 T; 0 other;

Query Match 80.0%; Score 16.8; DB 23; Length 3939;  
Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGGCACCATGTGATCCCCGA 20

||||| |||||||||

Db 3790 CGGCACCATGTGATCCCCCA 3809

Db

## RESULT 8

## RESULT 7

AAS94171  
ID AAS94171 standard; cDNA; 3939 BP.

XX AC AAS94171;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #29975.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG29984.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity

XX PS Claim 1; SEQ ID No 29975; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity  
XX CC and to produce other types of data and products dependent on DNA and  
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human  
XX CC diagnostic coding sequences of the invention.  
XX CC Note: The sequence data for this patent did not appear in the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3939 BP; 933 A; 1072 C; 1040 G; 894 T; 0 other;

Query Match 80.0%; Score 16.8; DB 23; Length 3939;  
Best Local Similarity 90.0%; Pred. No. 1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CGGCACCATGTGATCCCCGA 20

||||| |||||||||

Db 3790 CGGCACCATGTGATCCCCCA 3809

Db

```

AAI90923/c
ID AAI90923 standard; cDNA; 410 BP.
XX AC
XX AAI90923;
XX DT
XX DT (first entry)
XX DE
XX Human polynucleotide SEQ ID NO 10983.
XX KW
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation; ss.
XX OS
XX Homo sapiens.
XX PN
XX WO200164835-A2.
XX PD
XX 07-SEP-2001.
XX PF
XX 26-FEB-2001; 2001WO-US04927.
XX PR
XX 28-FEB-2000; 2000US-0515126.
XX PR
XX 18-MAY-2000; 2000US-0577409.
XX PA
XX (HYSE-) HYSEQ INC.
XX PI
XX Tang YT; Liu C, Drmanac RT;
XX PI
XX WPI; 2001-514838/56.
XX DR
XX P-PSDB; AAO10992.
XX PT
XX Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders.
XX PS
XX Claim 1; SEQ ID NO 10983; 1399pp + Sequence Listing; English.
XX CC
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ
XX Sequence 410 BP; 101 A; 81 C; 136 G; 92 T; 0 other;

Query Match      78.1%; Score 16.4; DB 22; Length 410;
Best Local Similarity 94.4%; Pred. No. 1.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CGGCACCATGTGATCCCC 18
        |||||
Db      228 CGGCACCATGTGATCCCC 211

RESULT 9
ID AAK84366/c
XX ID AAK84366 standard; DNA; 4829 BP.
XX AC
XX AAK84366;
XX DT
XX 07-NOV-2001 (first entry)
XX DE
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39178.

```

```

XX KW
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS
XX Homo sapiens.
XX PN
XX WO200157182-A2.
XX PD
XX 09-AUG-2001.
XX PF
XX 17-JAN-2001; 2001WO-US01354.
XX PR
XX 31-JAN-2000; 2000US-0179065.
XX PR
XX 04-FEB-2000; 2000US-0180628.
XX PR
XX 24-FEB-2000; 2000US-0184664.
XX PR
XX 02-MAR-2000; 2000US-0186350.
XX PR
XX 16-MAR-2000; 2000US-0189874.
XX PR
XX 17-MAR-2000; 2000US-0190076.
XX PR
XX 18-APR-2000; 2000US-0198123.
XX PR
XX 19-MAY-2000; 2000US-0205515.
XX PR
XX 07-JUN-2000; 2000US-0209467.
XX PR
XX 28-JUN-2000; 2000US-0214886.
XX PR
XX 30-JUN-2000; 2000US-0215135.
XX PR
XX 07-JUL-2000; 2000US-0216647.
XX PR
XX 07-JUL-2000; 2000US-0216880.
XX PR
XX 11-JUL-2000; 2000US-0217487.
XX PR
XX 11-JUL-2000; 2000US-0217496.
XX PR
XX 14-JUL-2000; 2000US-0218290.
XX PR
XX 26-JUL-2000; 2000US-0220963.
XX PR
XX 26-JUL-2000; 2000US-0220964.
XX PR
XX 14-AUG-2000; 2000US-0224518.
XX PR
XX 14-AUG-2000; 2000US-0224519.
XX PR
XX 14-AUG-2000; 2000US-0225213.
XX PR
XX 14-AUG-2000; 2000US-0225214.
XX PR
XX 14-AUG-2000; 2000US-0225266.
XX PR
XX 14-AUG-2000; 2000US-0225267.
XX PR
XX 14-AUG-2000; 2000US-0225268.
XX PR
XX 14-AUG-2000; 2000US-0225270.
XX PR
XX 14-AUG-2000; 2000US-0225447.
XX PR
XX 14-AUG-2000; 2000US-0225757.
XX PR
XX 14-AUG-2000; 2000US-0225758.
XX PR
XX 14-AUG-2000; 2000US-0225759.
XX PR
XX 18-AUG-2000; 2000US-0226279.
XX PR
XX 22-AUG-2000; 2000US-0226681.
XX PR
XX 22-AUG-2000; 2000US-0226688.
XX PR
XX 22-AUG-2000; 2000US-0227182.
XX PR
XX 23-AUG-2000; 2000US-0227009.
XX PR
XX 30-AUG-2000; 2000US-0228924.
XX PR
XX 01-SEP-2000; 2000US-0229287.
XX PR
XX 01-SEP-2000; 2000US-0229343.
XX PR
XX 01-SEP-2000; 2000US-0229344.
XX PR
XX 01-SEP-2000; 2000US-0229345.
XX PR
XX 05-SEP-2000; 2000US-0229509.
XX PR
XX 05-SEP-2000; 2000US-0229513.
XX PR
XX 06-SEP-2000; 2000US-0230437.
XX PR
XX 06-SEP-2000; 2000US-0230438.
XX PR
XX 08-SEP-2000; 2000US-0231242.
XX PR
XX 08-SEP-2000; 2000US-0231243.
XX PR
XX 08-SEP-2000; 2000US-0231244.
XX PR
XX 08-SEP-2000; 2000US-0231413.
XX PR
XX 08-SEP-2000; 2000US-0231414.
XX PR
XX 08-SEP-2000; 2000US-0232080.
XX PR
XX 08-SEP-2000; 2000US-0232081.
XX PR
XX 12-SEP-2000; 2000US-0231968.
XX PR
XX 14-SEP-2000; 2000US-0232397.
XX PR
XX 14-SEP-2000; 2000US-0232398.
XX PR
XX 14-SEP-2000; 2000US-0232399.
XX PR
XX 14-SEP-2000; 2000US-0232400.
XX PR
XX 14-SEP-2000; 2000US-0232401.
XX PR
XX 14-SEP-2000; 2000US-0233063.
XX PR
XX 14-SEP-2000; 2000US-0233064.
XX PR
XX 14-SEP-2000; 2000US-0233065.
XX PR
XX 21-SEP-2000; 2000US-02334223.

```

21-SEP-2000; 2000US-0234274.  
25-SEP-2000; 2000US-0234997.  
25-SEP-2000; 2000US-0234998.  
26-SEP-2000; 2000US-0235484.  
27-SEP-2000; 2000US-0235834.  
27-SEP-2000; 2000US-0235836.  
29-SEP-2000; 2000US-0236327.  
29-SEP-2000; 2000US-0236367.  
29-SEP-2000; 2000US-0236368.  
29-SEP-2000; 2000US-0236369.  
29-SEP-2000; 2000US-0236370.  
02-OCT-2000; 2000US-0236802.  
02-OCT-2000; 2000US-0237037.  
02-OCT-2000; 2000US-0237038.  
02-OCT-2000; 2000US-0237039.  
02-OCT-2000; 2000US-0237040.  
13-OCT-2000; 2000US-0239935.  
13-OCT-2000; 2000US-0239937.  
20-OCT-2000; 2000US-0240960.  
20-OCT-2000; 2000US-0241221.  
20-OCT-2000; 2000US-0241785.  
20-OCT-2000; 2000US-0241786.  
20-OCT-2000; 2000US-0241787.  
20-OCT-2000; 2000US-0241808.  
20-OCT-2000; 2000US-0241809.  
20-OCT-2000; 2000US-0241826.  
01-NOV-2000; 2000US-0244617.  
08-NOV-2000; 2000US-0246474.  
08-NOV-2000; 2000US-0246475.  
08-NOV-2000; 2000US-0246476.  
08-NOV-2000; 2000US-0246477.  
08-NOV-2000; 2000US-0246478.  
08-NOV-2000; 2000US-0246523.  
08-NOV-2000; 2000US-0246524.  
08-NOV-2000; 2000US-0246525.  
08-NOV-2000; 2000US-0246526.  
08-NOV-2000; 2000US-0246527.  
08-NOV-2000; 2000US-0246528.  
08-NOV-2000; 2000US-0246532.  
08-NOV-2000; 2000US-0246609.  
08-NOV-2000; 2000US-0246610.  
08-NOV-2000; 2000US-0246611.  
08-NOV-2000; 2000US-0246613.  
17-NOV-2000; 2000US-0249207.  
17-NOV-2000; 2000US-0249208.  
17-NOV-2000; 2000US-0249209.  
17-NOV-2000; 2000US-0249210.  
17-NOV-2000; 2000US-0249211.  
17-NOV-2000; 2000US-0249212.  
17-NOV-2000; 2000US-0249213.  
17-NOV-2000; 2000US-0249214.  
17-NOV-2000; 2000US-0249215.  
17-NOV-2000; 2000US-0249216.  
17-NOV-2000; 2000US-0249217.  
17-NOV-2000; 2000US-0249218.  
17-NOV-2000; 2000US-0249244.  
17-NOV-2000; 2000US-0249245.  
17-NOV-2000; 2000US-0249246.  
17-NOV-2000; 2000US-0249265.  
17-NOV-2000; 2000US-0249297.  
17-NOV-2000; 2000US-0249299.  
17-NOV-2000; 2000US-0249300.  
01-DEC-2000; 2000US-0250160.  
01-DEC-2000; 2000US-0250391.  
05-DEC-2000; 2000US-0251030.  
05-DEC-2000; 2000US-0251988.  
05-DEC-2000; 2000US-0256719.  
06-DEC-2000; 2000US-0251479.  
08-DEC-2000; 2000US-0251856.  
08-DEC-2000; 2000US-0251868.  
08-DEC-2000; 2000US-0251869.  
08-DEC-2000; 2000US-0251989.  
08-DEC-2000; 2000US-0251990.  
11-DEC-2000; 2000US-0254097.  
05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-483426/52.  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis.  
Disclosure; SEQ ID NO 39178; 3071pp + Sequence Listing; English.  
AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/hematopoietic-related diseases, especially cancers and cancer metastases of hematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/hematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.  
SQ Sequence 4829 BP; 922 A; 1179 C; 1681 G; 1047 T; 0 other;  
Query Match 78.1%; Score 16.4; DB 22; Length 4829;  
Best Local Similarity 94.4%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCC 18  
|||||||  
DB 1023 CGGCACCATGTGATCCCC 1006  
RESULT 10  
ABA70793  
ID ABA70793 standard; DNA; 97 BP.  
XX  
AC ABA70793;  
XX  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #19098..  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
OS Homo sapiens.  
XX WO200157277-A2;  
XX  
PN 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PR

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human fetal liver  
XX  
XX Claim 4; SEQ ID NO 19098; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX fetal liver. The present sequence is a single exon nucleic acid  
XX probe of the invention.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 97 BP; 22 A; 20 C; 29 G; 26 T; 0 other;  
SQ  
Query Match 77.1%; Score 16.2; DB 22; Length 97;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCCGAG 21  
Db 72 CTGCACCATGTGATCTCTGGAG 92  
RESULT 11  
AAK19046  
ID AAK19046 standard; DNA; 97 BP.  
XX  
XX AAK19046;  
XX  
XX 05-NOV-2001 (first entry)  
XX Human brain expressed single exon probe SEQ ID NO: 19037.  
XX Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains  
XX  
XX Example 4; SEQ ID NO: 19037; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
XX Sequence: 97 BP; 22 A; 20 C; 29 G; 26 T; 0 other;  
SQ  
Query Match 77.1%; Score 16.2; DB 22; Length 97;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCCGAG 21  
Db 72 CTGCACCATGTGATCTCTGGAG 92  
RESULT 12  
AAK44999  
ID AAK44999 standard; DNA; 97 BP.  
XX  
XX AAK44999;  
XX  
XX 06-NOV-2001 (first entry)  
XX Human bone marrow expressed single exon probe SEQ ID NO: 19556.  
XX Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX Homo sapiens.  
XX  
XX WO200157276-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00668.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow  
XX  
XX Example 4; SEQ ID NO: 19556; 658pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX the probes of the invention.  
XX  
XX Sequence 97 BP; 22 A; 20 C; 29 G; 26 T; 0 other;  
SQ  
Query Match 77.1%; Score 16.2; DB 22; Length 97;  
Best Local Similarity 85.7%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCCGAG 21  
Db 72 CTGCACCATGTGATCTCTGGAG 92

U  
U  
.

comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemangioendothelioma, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe open reading frame of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 97 BP; 22 A; 20 C; 29 G; 26 T; 0 other;

Query Match 77.1%; Score 16.2; DB 24; Length 97;

Best Local Similarity 85.7%; Pred. No. 1.6e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGGCACCATGTGATCCCGAG 21

Db 72 CTGCACCATGTGATCCTGGAG 92

RESULT 15

ABA58192

ID ABA58192 standard; DNA; 460 BP.

XX

AC ABA58192;

XX

DT 01-FEB-2002 (first entry)

XX

DE Human foetal liver single exon nucleic acid probe #6497.

XX

KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX

OS Homo sapiens.

XX

PN WO200157277-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00669.

XX

PR 04-FEB-2000; 2000US-0180312.

PR

26-MAY-2000; 2000US-0207456.

PR

30-JUN-2000; 2000US-0608408.

PR

03-AUG-2000; 2000US-0632366.

PR

21-SEP-2000; 2000US-0234687.

PR

27-SEP-2000; 2000US-0236359.

PR

04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DB;

XX

DR WPI; 2001-483447/52.

XX

PT Human genome-derived single exon nucleic acid probes useful for

XX

analyzing gene expression in human fetal liver.

PS

Claim 1; SEQ ID NO 6497; 639pp + sequence listing; English.

XX

CC The invention relates to a single exon nucleic acid probe for

CC

measuring human gene expression in a sample derived from human foetal

CC

liver. The single exon nucleic acid probes may be used for predicting,

CC

measuring and displaying gene expression in samples derived from human

CC

fetal liver. The present sequence is a single exon nucleic acid

CC

probe of the invention.

CC

Note: The sequence data for this patent did not form part of the

CC

printed specification, but was obtained in electronic format directly

CC

from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 460 BP; 138 A; 106 C; 100 G; 116 T; 0 other;

Query Match

Best Local Similarity 77.1%; Score 16.2; DB 22; Length 460;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 CGGCACCATGTGATCCCGAG 21

Db 372 CTGCACCATGTGATCCTGGAG 392

Search completed: July 8, 2003, 02:18:56  
Job time : 132.941 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	15.8	75.2	504	4	US-09-624-390-3	Sequence 3, Appli
2	15.8	75.2	1107	4	US-09-624-390-1	Sequence 1, Appli
3	15.2	72.4	1679	3	US-08-678-882-1	Sequence 1, Appli
4	15.2	72.4	1932	2	US-08-967-364-6	Sequence 6, Appli
5	15.2	72.4	1932	3	US-09-368-408-6	Sequence 6, Appli
6	15.2	72.4	1975	4	US-09-328-571A-12	Sequence 12, Appli
7	15.2	72.4	1977	1	US-08-273-255-17	Sequence 17, Appli
8	15.2	72.4	1977	5	PCT-US95-08365-17	Sequence 17, Appli
9	15.2	72.4	1989	2	US-08-792-035-1	Sequence 1, Appli
10	15.2	72.4	4214	4	US-09-221-017B-293	Sequence 293, App
11	15.2	72.4	35060	3	US-08-814-095-7	Sequence 7, Appli
12	15.2	72.4	152331	3	US-09-128-155-16	Sequence 16, Appli
13	15.2	72.4	176373	3	US-09-128-155-17	Sequence 17, Appli
14	15.2	72.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
15	15.2	72.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
16	14.8	70.5	529	4	US-08-998-416-26	Sequence 26, Appli
17	14.8	70.5	1432	4	US-09-183-861-73	Sequence 73, Appli
18	14.8	70.5	1432	4	US-09-022-765-73	Sequence 73, Appli
19	14.8	70.5	1912	1	US-08-270-013B-1	Sequence 73, Appli
20	14.8	70.5	1912	1	US-08-838-418-1	Sequence 1, Appli
21	14.8	70.5	3506	3	US-09-091-899-9	Sequence 1, Appli
22	14.6	69.5	32	1	US-08-465-687A-7	Sequence 7, Appli
23	14.6	69.5	32	3	US-09-030-970-7	Sequence 7, Appli
24	14.6	69.5	1353	4	US-09-724-864-32	Sequence 32, Appli
25	14.6	69.5	1542	5	PCT-US96-12345-4	Sequence 4, Appli
26	14.6	69.5	1552	5	PCT-US96-12345-1	Sequence 1, Appli
27	14.6	69.5	1555	5	PCT-US96-12345-2	Sequence 2, Appli
28	14.6	69.5	1555	5	PCT-US96-12345-2	Sequence 2, Appli

```

: GENERAL INFORMATION:
: APPLICANT: VERSECK, STEFAN
: APPLICANT: KULA, MARIA-REGINA
: APPLICANT: BONMARIUS, ANDREAS
: APPLICANT: DRAUZ, KARLHEINZ
: TITLE OF INVENTION: N-ACETYL AMINO ACID RACEMASE
: FILE REFERENCE: 192535US0
: CURRENT APPLICATION NUMBER: US/09/624,390
: CURRENT FILING DATE: 2000-07-27
: PRIOR APPLICATION NUMBER: DE 19935268.2
: PRIOR FILING DATE: 1999-07-27
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1

```

;; LENGTH: 1107  
;; TYPE: DNA  
;; ORGANISM: Amycolatopsis orientalis  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1)..(1107)  
US-09-624-390-1

Query Match 75.2%; Score 15.8; DB 4; Length 1107;  
Best Local Similarity 89.5%; Pred. No. 42;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGACCATGTGATCCCG 19  
DB 202 CGGAACCATGTGATCCCG 220

RESULT 3  
US-08-676-882-1/c  
; Sequence 1, Application US/08676882  
; Patent No. 6100241  
; GENERAL INFORMATION:  
; APPLICANT: Kok, Jacobus Johannes  
; APPLICANT: van den Boogaart, Paul  
; APPLICANT: Vermeulen, Arnoldus Nicolaas  
; TITLE OF INVENTION: Coccidiosis poultry vaccine  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Akzo No. 6100241el Patent Department  
; STREET: 1300 Piccard Drive, Suite 206  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,882  
FILING DATE: 03-JUL-1996  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Gormley, Mary E.  
REGISTRATION NUMBER: 34,409  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 258-5200  
TELEFAX: (301) 977-0847

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1679 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:

ORGANISM: Eimeria acervulina  
DEVELOPMENTAL STAGE: Schizont  
IMMEDIATE SOURCE:  
CLONE: EASC2\_1

FEATURE:  
NAME/KEY: CDS

LOCATION: 280..1269  
OTHER INFORMATION: /function- "Eimeria lactate  
OTHER INFORMATION: dehydrogenase"

NAME/KEY: misc feature  
LOCATION: 1..51

OTHER INFORMATION: /label- pbluescriptII  
FEATURE:

;; NAME/KEY: misc feature  
;; LOCATION: 1624..1679  
;; OTHER INFORMATION: /label- pbluescriptII  
;; FEATURE:

;; NAME/KEY: misc feature  
;; LOCATION: 45..54  
;; OTHER INFORMATION: /label- EcoRI-linker  
;; FEATURE:

;; NAME/KEY: misc feature  
;; LOCATION: 1621..1630  
;; OTHER INFORMATION: /label- EcoRI-linker  
US-08-676-882-1

Query Match 72.4%; Score 15.2; DB 3; Length 1679;  
Best Local Similarity 85.0%; Pred. No. 86;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21  
DB 857 GGCACCATGTGATCCCGTG 838

RESULT 4  
US-08-967-364-6/c  
; Sequence 6, Application US/08967364  
; Patent No. 5989859  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Shah, Purvi  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: VESICLE TRAFFICKING PROTEINS  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Dr.  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/967,364  
FILING DATE: No. 5989859ember 7, 1997  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Certone, Michael C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0417 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1932 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: 3086794  
CLONE: HEAONOT03  
US-08-967-364-6

Query Match 72.4%; Score 15.2; DB 2; Length 1932;  
Best Local Similarity 85.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;



## RESULT 8

PCT-US95-08565-17/c  
Sequence 17, Application PC/TUS9508565.  
GENERAL INFORMATION:  
APPLICANT: Cashmore, Anthony R.  
APPLICANT: Ahmad, Margaret  
APPLICANT: Lin, Chentao  
TITLE OF INVENTION: Blue Light Photoreceptors and Methods of  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/08565  
FILING DATE:  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/272,255  
FILING DATE: 08-JUL-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: UPN-1795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1977 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
PCT-US95-08565-17

Query Match 72.4%; Score 15.2; DB 5; Length 1977;  
Best Local Similarity 85.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCACCATGTGATCCCGG 21  
DB 1321 GCACCATGTGATCTCGG 1302

## RESULT 9

US-08-792-055-1/c  
Sequence 1, Application US/08792055  
Patent No. 5853980  
GENERAL INFORMATION:  
APPLICANT: Rollin, Pierre E.  
APPLICANT: Elliott, Luanne  
APPLICANT: Ksiazek, Thomas G.  
APPLICANT: Nichol, Stuart T.  
APPLICANT: Morzunov, Sergey  
APPLICANT: Ravkov, Eugene  
TITLE OF INVENTION: The Black Creek Canal Hantavirus and  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
STREET: 127 Peachtree Street, N.E., Suite 1200  
CITY: Atlanta  
STATE: Georgia

COUNTRY: USA  
ZIP: 30303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/792,055  
FILING DATE: 03-FEB-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/390,361  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 1414,622  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 688-0770  
TELEFAX: (404) 688-9880

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1989 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-792-055-1

Query Match 72.4%; Score 15.2; DB 2; Length 1989;  
Best Local Similarity 85.0%; Pred. No. 87;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCACCATGTGATCCCGG 21  
DB 1250 GCATCCATGTCATCCCGG 1231

## RESULT 10

US-09-221-017B-293/c  
Sequence 293, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PP2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: MONROY, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 293:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4214 base pairs  
TYPE: nucleic acid  
STRADEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...4214  
US-09-221-017B-293

Query Match 72.4%; Score 15.2; DB 4; Length 4214;  
Best Local Similarity 85.0%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCGA 20  
||||| ||||||| |||  
DB 3755 CGGCTCATGTGATCGAGA 3736

RESULT 11  
US-08-814-095-7  
Sequence 7, Application US/08814095  
Patent No. 6025183  
GENERAL INFORMATION:  
APPLICANT: Soreq, Hermona  
APPLICANT: Zakut, Haim  
APPLICANT: Shani, Moshe  
TITLE OF INVENTION: TRANSGENIC ANIMAL ASSAY SYSTEM FOR  
TITLE OF INVENTION: ANTI-CHOLINESTERASE SUBSTANCES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: KOHN & ASSOCIATES  
STREET: 30500 No. 6025183thwestern Highway, Suite 410  
CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: U.S.  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/814,095  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Montgomery, Ilene N.  
REGISTRATION NUMBER: 38,972  
REFERENCE/DOCKET NUMBER: 2391.00066  
TELEPHONE: (248) 539-5050  
TELEFAX: (248) 539-5055  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35060 base pairs  
TYPE: nucleic acid

STRADEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Cosmid including ACHE  
DESCRIPTION: promotor, ACHE gene and ARS gene"  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 7q22  
FEATURE:  
NAME/KEY: promotor  
LOCATION: 4089...22464  
OTHER INFORMATION: /function= "ACHE Promotor"  
OTHER INFORMATION: /standard\_name= "ACHE Promotor"  
FEATURE:  
NAME/KEY: exon  
LOCATION: 22465...22537  
OTHER INFORMATION: /function= "non-translated"  
OTHER INFORMATION: /gene= "ACHE"  
OTHER INFORMATION: /number= 1  
FEATURE:  
NAME/KEY: exon  
LOCATION: 24090...25177  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /function= "(translation start:  
OTHER INFORMATION: 24110)"  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "ACHE"  
OTHER INFORMATION: /number= 2  
FEATURE:  
NAME/KEY: exon  
LOCATION: 25524...26009  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "ACHE"  
OTHER INFORMATION: /number= 3  
FEATURE:  
NAME/KEY: exon  
LOCATION: 27005...27274  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "ACHE"  
OTHER INFORMATION: /number= 4  
FEATURE:  
NAME/KEY: exon  
LOCATION: 27255...28007  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "ACHE"  
OTHER INFORMATION: /number= 5  
FEATURE:  
NAME/KEY: terminator  
LOCATION: 27385...27387  
FEATURE:  
NAME/KEY: exon  
LOCATION: 28008...28129  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /evidence= EXPERIMENTAL  
OTHER INFORMATION: /gene= "ACHE"  
OTHER INFORMATION: /number= 6  
FEATURE:  
NAME/KEY: terminator  
LOCATION: 28129...28131  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (34528...34895)  
OTHER INFORMATION: /function= "arsenite resistance  
OTHER INFORMATION: gene"  
OTHER INFORMATION: /gene= "AR"  
OTHER INFORMATION: /number= 1  
FEATURE:

NAME/KEY: exon  
LOCATION: complement (34092..34358)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 2  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (33779..33963)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 3  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (33493..33591)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 4  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (33297..33408)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 5  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (32959..33094)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 6  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (32569..32628)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 7  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (32386..32468)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 8  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (31894..32080)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 9  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (31363..31534)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 10  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (31131..31284)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 11  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (30816..31011)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 12  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (30470..30626)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 13  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (30187..30274)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 14  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (29945..30073)  
OTHER INFORMATION: /gene="AR"  
OTHER INFORMATION: /number= 15  
FEATURE:  
NAME/KEY: exon  
LOCATION: complement (29664..29856)  
OTHER INFORMATION: /gene="ARs"

OTHER INFORMATION: /number= 16  
US-08-814-095-7  
Query Match 72.4%; Score 15.2; DB 3; Length 35060;  
Best Local Similarity 85.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 GGCACCATGTGATCCCGAG 21  
||||| ||| |||||  
DB 22585 GGCACCGTGGCTCCCGAG 22604  
RESULT 12  
US-09-128-155-16/c  
; Sequence 16, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; CURRENT FILING DATE: 1998-08-03  
; EARLIER APPLICATION NUMBER: US 60/091,650  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: US 60/054,646  
; EARLIER FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 16  
; LENGTH: 152331  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(152331)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-128-155-16  
Query Match 72.4%; Score 15.2; DB 3; Length 152331;  
Best Local Similarity 85.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 GGCACCATGTGATCCCGAG 21  
||||| ||||| |||||  
DB 65542 GGCACCGTGGTATCCCGAG 65523  
RESULT 13  
US-09-128-155-17  
; Sequence 17, Application US/09128155  
; Patent No. 6117654  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Yang  
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY  
; TITLE OF INVENTION: AND USES THEREOF  
; FILE REFERENCE: 09404/052001  
; CURRENT APPLICATION NUMBER: US/09/128,155  
; CURRENT FILING DATE: 1998-08-03  
; EARLIER APPLICATION NUMBER: US 60/091,650  
; EARLIER FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: US 60/054,646  
; EARLIER FILING DATE: 1997-08-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 176373  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(176373)  
; OTHER INFORMATION: n = A,T,C or G

US-09-128-155-17

Query Match 72.4%; Score 15.2; DB 3; Length 176373;  
Best Local Similarity 85.0%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GGCACCATGTGATCCCCGAG 21  
||||| ||||| ||||| |||||  
Db 69408 GGCACCATGTGATCCCCAGTG 69427

## RESULT 14

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103.840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 72.4%; Score 15.2; DB 4; Length 4403765;  
Best Local Similarity 85.0%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGGCACCATGTGATCCCCGA 20  
||||| ||||| ||||| |||||  
Db 3360524 CGGCACCATGTGCTCGCCGA 3360505

## RESULT 15

US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103.840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 72.4%; Score 15.2; DB 4; Length 4411529;  
Best Local Similarity 85.0%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CGGCACCATGTGATCCCCGA 20  
||||| ||||| ||||| |||||  
Db 3366192 CGGCACCATGTGCTCGCCGA 3366173

Search completed: July 8, 2003, 09:31:44  
Job time : 45.0402 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 121.288 seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-10

Perfect score: 21

Sequence: 1 cggcaccatgtatcccgag 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.2	77.1	97	10	US-09-864-761-27728
2	16.2	77.1	458	9	US-09-918-995-25291
3	16.2	77.1	460	10	US-09-864-761-11096
4	16.2	77.1	464	9	US-09-918-995-22699
5	16.2	77.1	1150	9	US-10-239-420-13
6	16.2	77.1	2893	9	US-10-017-273A-4
7	16.2	77.1	2893	10	US-09-905-846-1
8	16.2	77.1	2975	9	US-10-017-273A-5
9	16.2	77.1	2975	10	US-09-905-846-5
10	15.8	75.2	600	9	US-10-156-761-6591
11	15.8	75.2	1107	10	US-09-973-765-1
12	15.8	75.2	1107	10	US-09-973-712-1
13	15.8	75.2	9025608	9	US-10-156-761-1
14	15.4	73.3	970	10	US-09-764-864-284
15	15.2	72.4	356	9	US-09-933-797-380
16	15.2	72.4	445	10	US-09-867-550-63
17	15.2	72.4	494	9	US-09-918-995-31268
18	15.2	72.4	569	10	US-09-764-877-903
19	15.2	72.4	831	10	US-09-867-701-10019

c 20	15.2	72.4	888	10	US-09-822-830A-590
21	15.2	72.4	915	9	US-10-156-761-2013
c 22	15.2	72.4	1152	9	US-10-092-154-1517
c 23	15.2	72.4	1152	10	US-09-764-847-1517
c 24	15.2	72.4	1411	9	US-10-037-270-535
c 25	15.2	72.4	1655	9	US-10-092-154-1518
c 26	15.2	72.4	1655	10	US-09-764-847-1518
c 27	15.2	72.4	1738	10	US-09-925-297-317
c 28	15.2	72.4	2031	12	US-10-044-090-468
c 29	15.2	72.4	2564	10	US-09-880-107-3753
c 30	15.2	72.4	152331	9	US-10-095-407-16
31	15.2	72.4	176373	9	US-10-095-407-17
32	15.2	72.4	180216	10	US-09-835-232-6
c 33	14.8	70.5	240	10	US-09-923-876-2432
c 34	14.8	70.5	324	10	US-09-974-300-942
c 35	14.8	70.5	457	10	US-09-783-590-3063
c 36	14.8	70.5	491	9	US-10-079-623-81
37	14.8	70.5	754	9	US-10-106-698-1121
c 38	14.8	70.5	1432	9	US-09-991-496-73
c 39	14.8	70.5	1432	10	US-09-874-923-73
c 40	14.8	70.5	1527	9	US-10-079-854-319
c 41	14.8	70.5	1527	10	US-09-764-878-319
c 42	14.8	70.5	1688	9	US-09-991-496-105
c 43	14.8	70.5	1688	10	US-09-874-923-105
c 44	14.8	70.5	2047	10	US-09-822-830A-587
c 45	14.8	70.5	3129	9	US-10-161-803-24

ALIGNMENTS

RESULT 1

US-09-864-761-27728  
; Sequence 27728; Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David K.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864.761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30

; Sequence 22699, Application US/09918995

Db 557 CGGGCCACGTGATCCCTGAG 577  
 ||||| ||||| ||||| ||||| |||||  
 RESULT 8  
 US-10-017-273A-5  
 ; Sequence 5, Application US/10017273A  
 ; Publication No. US20030119714A1  
 ; GENERAL INFORMATION:

APPLICANT: Pfizer Inc.  
APPLICANT: Naylor, Alasdair M.  
APPLICANT: Van Der Graaf, Pieter H  
APPLICANT: Wayman, Christopher P.  
TITLE OF INVENTION: Treatment of Male Sexual Dysfunction  
FILE REFERENCE: PC22013  
CURRENT APPLICATION NUMBER: US/10/017,273A  
CURRENT FILING DATE: 2001-12-12  
PRIOR FILING DATE: 2001-01-31  
PRIOR APPLICATION NUMBER: GB 0030647.2  
PRIOR FILING DATE: 2000-12-15  
PRIOR APPLICATION NUMBER: GB 0108730.3  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: GB 0120679.6  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 09/905,846  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/291,722  
PRIOR FILING DATE: 2001-05-17  
PRIOR APPLICATION NUMBER: US 09/895,367  
PRIOR FILING DATE: 2001-06-29  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 5  
LENGTH: 2975  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-017-273A-5

Query Match 77.1%; Score 16.2; DB 9; Length 2975;  
Best Local Similarity 85.7%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCCGAG 21  
||||| ||| ||||| ||| |||  
DB 622 CGGCGCCAGTGTATCCCTGAG 642

RESULT 9  
US-09-905-846-5  
Sequence 5, Application US/09905846  
Patent No. US20020102707A1  
GENERAL INFORMATION:  
APPLICANT: Ian Dennis Harrow  
APPLICANT: Peter Stacey  
APPLICANT: Roderick Thomas Walsh  
TITLE OF INVENTION: Compounds for the treatment of sexual dysfunction  
FILE REFERENCE: PCS10926APME  
CURRENT APPLICATION NUMBER: US/09/905,846  
CURRENT FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: 0017387.2  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/220,908  
PRIOR FILING DATE: 2000-07-26  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 2975  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-905-846-5

Query Match 77.1%; Score 16.2; DB 10; Length 2975;  
Best Local Similarity 85.7%; Pred. No. 72;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCCGAG 21  
||||| ||| ||||| ||| |||  
DB 622 CGGCGCCAGTGTATCCCTGAG 642

RESULT 10

US-10-156-761-6591  
Sequence 6591, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: HORIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 6591  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(600)  
US-10-156-761-6591

Query Match 75.2%; Score 15.8; DB 9; Length 600;  
Best Local Similarity 89.5%; Pred. No. 11e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 3 GCACCATGTGATCCCCGAG 21  
||||| ||| ||||| |||  
DB 348 GCACCATGTGATCCCCGAG 366

RESULT 11  
US-09-973-765-1  
Sequence 1, Application US/09973765  
Patent No. US20020090684A1  
GENERAL INFORMATION:  
APPLICANT: BOMMARIUS, ANDREAS  
APPLICANT: DRAUZ, KARLHEINZ  
APPLICANT: VERSECK, STEFAN  
TITLE OF INVENTION: PROCESS FOR THE PRODUCTION OF AMINO ACIDS  
FILE REFERENCE: 214381US-10757-9350-0-X  
CURRENT APPLICATION NUMBER: US/09/973,765  
CURRENT FILING DATE: 2001-10-11  
PRIOR APPLICATION NUMBER: DE 100 50 123.0  
PRIOR FILING DATE: 2000-10-11  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1107  
TYPE: DNA  
ORGANISM: Amycolatopsis orientalis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1107)  
OTHER INFORMATION:  
US-09-973-765-1

Query Match 75.2%; Score 15.8; DB 10; Length 1107;  
Best Local Similarity 89.5%; Pred. No. 1.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCCG 19  
||||| ||| ||||| ||| |||  
DB 202 CGGAACCATCTGATCCCCG 220

## RESULT 12

US-09-973-712-1  
; Sequence 1, Application US/09973712  
; Patent No. US20020106752A1

## GENERAL INFORMATION:

; APPLICANT: BOMMARIUS, ANDREAS  
; APPLICANT: DRAUZ, KARLHEINZ  
; APPLICANT: VERSECK, STEFAN

; APPLICANT: KULA, MARIA-REGINA

; TITLE OF INVENTION: ACETYL AMINO ACID RACEMASE FROM AMYCOLATOPSIS ORIENTALIS FOR RACE

; FILE REFERENCE: 214382USOX

; CURRENT APPLICATION NUMBER: US/09/973,712

; CURRENT FILING DATE: 2001-10-11

; PRIOR APPLICATION NUMBER: DE 10050124.9

; PRIOR FILING DATE: 2000-10-11

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1107

; TYPE: DNA

; ORGANISM: Amycolatopsis orientalis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1107)

; OTHER INFORMATION:

US-09-973-712-1

## Query Match

Best Local Similarity 75.2%; Score 15.8; DB 10; Length 1107;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGACCATGTGATCCCCG 19

DB 202 CGGAACCATGTGATCCCCG 220

## RESULT 13

US-10-156-761-1

; Sequence 1, Application US/10156761

; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 1

LENGTH: 9025608

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: misc.feature

LOCATION: (4187715)

OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

## Query Match

Best Local Similarity 75.2%; Score 15.8; DB 9; Length 9025608;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GCACCATGTGATCCCCGAG 21

||||||| ||| |||||||

Db 7913762 GCACCATGTGATCCCCGAG 7913780

## RESULT 14

US-09-764-864-284/c

; Sequence 284, Application US/09764864

; Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: ROSEN ET AL.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT223

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 284

LENGTH: 970

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-864-284

## Query Match

Best Local Similarity 73.3%; Score 15.4; DB 10; Length 970;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCC 18

Db 946 GGCACCATGTGATCCCC 930

## RESULT 15

US-09-933-797-380

; Sequence 380, Application US/09933797

; Patent No. US20020155119A1

GENERAL INFORMATION:

APPLICANT: ROBERT A. SIKES ET AL.

TITLE OF INVENTION: Isolation and Use of Fetal Urogenital

TITLE OF INVENTION: Sinus Expressed Sequences

FILE REFERENCE: 9901-007-999

CURRENT APPLICATION NUMBER: US/09/933,797

CURRENT FILING DATE: 2001-08-22

PRIOR APPLICATION NUMBER: US/09/482,933

PRIOR FILING DATE: 2000-01-14

PRIOR APPLICATION NUMBER: PCT/US99/10746

PRIOR FILING DATE: 1999-05/14

PRIOR APPLICATION NUMBER: 60/085,383

PRIOR FILING DATE: 1998-05-14

NUMBER OF SEQ ID NOS: 811

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 380

LENGTH: 356

TYPE: DNA

ORGANISM: Murine

US-09-933-797-380

## Query Match

Best Local Similarity 72.4%; Score 15.2; DB 9; Length 356;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCCGAG 21

Db 85 GGCACCATGTGATCTCCAAG 104

Search completed: July 9, 2003, 02:22:04

Job time : 132.338 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1064.65 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-10

Perfect score: 21

Sequence: 1 cggcaccatgtatcccccag 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_man.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	17.8	84.8	242	14	BQ990877 QGF21F15.
c 2	17.8	84.8	413	14	BQ980809 QGF12B10.
c 3	17.8	84.8	557	14	BQ010846 QGF14J23.
c 4	17.8	84.8	565	14	BQ983268 QGF18K14.
c 5	17.8	84.8	648	14	BQ868940 QGD39G03.Y
c 6	17.4	82.9	536	9	A1257845 LP06332.5

c 7	81.0	17	325	12	BQ055194
c 8	80.0	16.8	156	9	AV005955
c 9	80.0	16.8	167	9	AV132551
c 10	80.0	16.8	189	9	AV164385
c 11	80.0	16.8	200	10	BB067107
c 12	80.0	16.8	204	14	BQ417729
c 13	80.0	16.8	207	12	BG791763
c 14	80.0	16.8	212	10	BB185381
c 15	80.0	16.8	225	10	BB149886
c 16	80.0	16.8	227	9	AV222275
c 17	80.0	16.8	228	10	AV308313
c 18	80.0	16.8	228	10	AV381129
c 19	80.0	16.8	228	10	BB069093
c 20	80.0	16.8	230	9	AV088272
c 21	80.0	16.8	230	9	AV297618
c 22	80.0	16.8	230	10	BB425101
c 23	80.0	16.8	231	9	AV113705
c 24	80.0	16.8	232	9	AV086443
c 25	80.0	16.8	232	10	BB068381
c 26	80.0	16.8	232	10	BB135193
c 27	80.0	16.8	233	9	AV252958
c 28	80.0	16.8	233	9	AV282409
c 29	80.0	16.8	233	10	AV378061
c 30	80.0	16.8	234	10	BB480218
c 31	80.0	16.8	235	10	BB292828
c 32	80.0	16.8	235	10	BB362719
c 33	80.0	16.8	236	10	BB043974
c 34	80.0	16.8	236	10	BB097432
c 35	80.0	16.8	236	10	BB485095
c 36	80.0	16.8	238	10	BB546577
c 37	80.0	16.8	239	10	BB558444
c 38	80.0	16.8	243	9	AV051230
c 39	80.0	16.8	243	10	AV363610
c 40	80.0	16.8	247	10	AV363901
c 41	80.0	16.8	250	9	AV140996
c 42	80.0	16.8	251	9	AV132828
c 43	80.0	16.8	251	10	AV309958
c 44	80.0	16.8	251	10	AV377283
c 45	80.0	16.8	252	10	BB142844

#### ALIGNMENTS

RESULT 1  
BQ990877/c  
LOCUS BQ990877.1  
DEFINITION QGF21F15.Yg.ab1 QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
ACCESSION BQ990877  
VERSION BQ990877.1 GI:22410412  
KEYWORDS EST.  
SOURCE Lactuca sativa.  
ORGANISM Lactuca sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.  
REFERENCE 1 (bases 1 to 242)  
AUTHORS Kosik A., Michelson R.W., Knapp S., Matvienko M., Rieseberg L., Lin H., van Damme M., Lavelle D., Chevallier P., Ziegler J., Ellison P., Kolkman J., Slabaugh M.S., Livingston K., Zhou Y., Lai Z., Church S., Jackson L. and Bradford K.  
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Alexander Kozik [R.W.Michelson@lab Department of Vegetable Crops, R.W.Michelson Lab University of California at Davis (UCD) Asmudson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@ucdavis.org [michelson@vegmail.ucdavis.edu]

belongs to contig QG\_CA\_Contig3186, see <http://cgpdb.ucdavis.edu/> for details.

Plate: QGF21 row: F column: 15.

#### FEATURES

source

1. 242  
Location/Qualifiers  
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db\_xref="taxon:4236"  
/clone="QG21F15"  
/clone\_lib="QG\_EFGHJ lettuce serriola"  
/lab\_host="E.coli"  
/note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>  
TAG\_TISSUE=leaves dark grow  
TAG\_SEQ=GCTAGTCGGG"

BASE COUNT 38 a 35 c 131 g 38 t

#### ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 242;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCGAG 21  
|||||  
Db 124 CGGCACCATGTGATCCCGAG 104  
|||||

#### RESULT 2

BQ980809/c  
LOCUS BQ980809 413 bp mRNA linear EST 21-AUG-2002  
DEFINITION QGE12B10, yg abl QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
ACCESSION BQ980809.1 GI:22398332  
VERSION BQ980809.1  
KEYWORDS EST.  
SOURCE Lactuca sativa.  
ORGANISM Lactuca sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

REFERENCE 1 (bases 1 to 413)  
AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.  
TITLE Lactuca sativa  
JOURNAL Lettuce and Sunflower ESTs from the Compositae Genome Project  
COMMENT Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmudson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

belongs to contig QG\_CA\_Contig3186, see <http://cgpdb.ucdavis.edu/> for details.  
Plate: QGE12 row: B column: 10.  
Location/Qualifiers  
1. 413  
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db\_xref="taxon:4236"  
/clone="QGE12B10"

#### FEATURES

source

1. 413  
Location/Qualifiers  
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db\_xref="taxon:4236"  
/clone="QGE12B10"

/clone\_lib="QG\_EFGHJ lettuce serriola"  
/lab\_host="E.coli"  
/note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>  
TAG\_TISSUE=chemical induction  
TAG\_SEQ=TGTAGCCGGG"

BASE COUNT 68 a 52 c 222 g 71 t

#### ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 413;  
Best Local Similarity 90.5%; Pred. No. 3.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CGGCACCATGTGATCCCGAG 21  
|||||  
Db 174 CGGCACCATGTGATCCCGAG 154  
|||||

#### RESULT 3

BQ010846/c  
LOCUS BQ010846 557 bp mRNA linear EST 22-AUG-2002  
DEFINITION QGJ14J23, yg abl QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
ACCESSION QGJ14J23.1 GI:22445241  
VERSION BQ010846.1  
KEYWORDS EST.  
SOURCE Lactuca sativa.  
ORGANISM Lactuca sativa  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae; Lactuca.

REFERENCE 1 (bases 1 to 557)

AUTHORS Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.  
TITLE Lettuce and Sunflower ESTs from the Compositae Genome Project  
COMMENT Unpublished (2002)  
Contact: Alexander Kozik [R.W.Michelmore]  
Department of Vegetable Crops, R.W.Michelmore Lab  
University of California at Davis (UCD)  
Asmudson Hall, UCD, Davis, CA 95616, USA  
Tel: 1-(530)-742-1742  
Fax: 1-(530)-752-9659  
Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu]

#### JOURNAL

COMMENT

#### FEATURES

source

1. 557  
Location/Qualifiers  
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db\_xref="taxon:4236"  
/clone="QGJ14J23"  
/clone\_lib="QG\_EFGHJ lettuce serriola"  
/lab\_host="E.coli"

/note="Vector: pBRCNDSFIAB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/> for details.  
Plate: QGJ14 row: J column: 23.  
Location/Qualifiers  
1. 557  
/organism="Lactuca sativa"  
/cultivar="L. serriola"  
/db\_xref="taxon:4236"  
/clone="QGJ14J23"  
/clone\_lib="QG\_EFGHJ lettuce serriola"  
/lab\_host="E.coli"

construction can be obtained at <http://cgpdb.ucdavis.edu/>  
 TAG\_LIB=QG\_EFGHJ lettuce serriola  
 TAG\_TISSUE=flowers pre-fertilized  
 TAG\_SEQ=GCTGACGGG"

BASE COUNT 132 a 82 c 202 g 141 t  
 ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 557;  
 Best Local Similarity 90.5%; Pred. No. 4.3e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCGAG 21  
 |||||  
 Db 64 CGGCACCATGTGATCTCCAG 44

## RESULT 4

BQ983268/c  
 LOCUS QGE18K14.yg.abl QG\_EFGHJ lettuce serriola Lactuca sativa cDNA clone  
 DEFINITION QGE18K14, mRNA sequence.

ACCESSION BQ983268  
 VERSION 1 GI:22400793  
 KEYWORDS EST.

## ORGANISM

Lactuca sativa.  
 Lactuca sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.

## REFERENCE

1 (bases 1 to 565)  
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
 Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://comgenomics.ucdavis.edu/  
 Unpublished (2002)

## JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Fax: 1-(530)-742-1742  
 Tel: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QG\_CA\_Contig3575, see <http://cgpdb.ucdavis.edu/>  
 for details.

Plate: QGE18 row: K column: 14.

## FEATURES

## source

1..565  
 Location/Qualifiers  
 /organism="Lactuca sativa"  
 /cultivar="L.serriola"  
 /db\_xref="taxon:4236"  
 /clone="QGE18K14"  
 /clone\_lib="QG\_EFGHJ lettuce serriola"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASFIAB; The library was constructed

from 10 different sources of RNA from a single genotype.  
 Separate cDNAs were generated using primers that  
 incorporated unique 5' and 3' tags to distinguish each  
 source of RNA. cDNAs were then pooled, size-fractionated,  
 directionally cloned into a custom medium-copy vector and  
 transformations made with four size classes to minimize  
 size bias. Details of each source of RNA and library  
 construction can be obtained at <http://cgpdb.ucdavis.edu/>  
 TAG\_LIB=QG\_EFGHJ lettuce serriola  
 TAG\_TISSUE=leaves dark grow  
 TAG\_SEQ=GCTGACGGG"

BASE COUNT 130 a 82 c 220 g 133 t  
 ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 565;  
 Best Local Similarity 90.5%; Pred. No. 4.3e+02;

## Matches

19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCGAG 21  
 |||||  
 Db 99 CGGCACCATGTGATCTCCAG 79

## RESULT 5

BQ868940/c  
 LOCUS QGD3G03.yg.abl QG\_ABCDI lettuce salinas Lactuca sativa cDNA clone  
 DEFINITION QGD3G03, mRNA sequence.

ACCESSION BQ868940  
 VERSION 1 GI:22254697  
 KEYWORDS EST.

## ORGANISM

Lactuca sativa.  
 Lactuca sativa  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;  
 Lactuca.

## REFERENCE

1 (bases 1 to 648)  
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,  
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison  
 P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,  
 Church, S., Jackson, L. and Bradford, K.  
 Lettuce and Sunflower ESTs from the Compositae Genome Project  
 http://comgenomics.ucdavis.edu/  
 Unpublished (2002)

## JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]  
 Department of Vegetable Crops, R.W.Michelmore Lab  
 University of California at Davis (UCD)  
 Asmundson Hall, UCD, Davis, CA 95616, USA  
 Tel: 1-(530)-742-1742  
 Fax: 1-(530)-752-9659  
 Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]  
 belongs to contig QGDB.ucdavis.edu/ for details.  
 Plate: QGD3 row: g column: 03.

## FEATURES

## source

1..648  
 Location/Qualifiers  
 /organism="Lactuca sativa"  
 /cultivar="Salinas"  
 /db\_xref="taxon:4236"  
 /clone="QGD3G03"  
 /clone\_lib="QG\_ABCDI lettuce salinas"  
 /lab\_host="E.coli"  
 /note="Vector: pBRCDNASFIAB; The library was constructed

from 10 different sources of RNA from a single genotype.  
 Separate cDNAs were generated using primers that  
 incorporated unique 5' and 3' tags to distinguish each  
 source of RNA. cDNAs were then pooled, size-fractionated,  
 directionally cloned into a custom medium-copy vector and  
 transformations made with four size classes to minimize  
 size bias. Details of each source of RNA and library  
 construction can be obtained at <http://cgpdb.ucdavis.edu/>  
 TAG\_LIB=QG\_ABCDI lettuce salinas  
 TAG\_TISSUE=leaves dark grow  
 TAG\_SEQ=GCTAGTCGGG"

BASE COUNT 155 a 140 c 213 g 140 t  
 ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 648;  
 Best Local Similarity 90.5%; Pred. No. 4.5e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGGCACCATGTGATCCCGAG 21  
 |||||  
 Db 64 CGGCACCATGTGATCTCCAG 44

RESULT 6  
 A1257845  
 LOCUS

A1257845

536 bp

mRNA

linear

EST 19-APR-2001

```

DEFINITION LP06332.5prime LP Drosophila melanogaster larval-early pupal pOT2
Drosophila melanogaster cDNA clone LP06332 5prime similar to
X54997: D.melanogaster P1 gene, mRNA sequence.
ACCESSION AI257845.1 GI:3865370
VERSION AI257845.1
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 536)
AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,
Lewis, S., and Rubin, G.M.
TITLE BCGP/HMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
Plate: 63 row: C column: 8
High quality sequence stop: 476.
FEATURES
    source
        1..536
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="LP06332"
            /clone_lib="LP Drosophila melanogaster larval-early pupal
            pOT2"
            /sex="male and female"
            /dev_stage="larvae-pupae"
            /lab_host="DH5-alpha"
            /note="Organ: whole body; Vector: pOT2; Site 1: ECORI;
            Site 2: XbaI; Sized fractionated cDNAs were directly
            ligated into pOT2. Plasmid cDNA library."
BASE COUNT 136 a 126 c 143 g 131 t
ORIGIN
    Query Match 82.9%; Score 17.4; DB 9; Length 536;
    Best Local Similarity 94.7%; Pred. No. 6.5e+02;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 GCACCATGTGATCCCGAG 21
Db 384 GCACCATGTGATCCCGAG 402
    |||||||
RESULT 7
BG055194/c
LOCUS BG055194 325 bp mRNA linear EST 25-JAN-2001
DEFINITION nad03e01.x1 NCI-CGAP Pr28 Homo sapiens cDNA clone IMAGE:3432289 3'
similar to contains Alu repetitive element; contains element MER22
repetitive element ;, mRNA sequence.
ACCESSION BG055194
VERSION BG055194.1 GI:12512671
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 325)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.

```

```

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
FEATURES
    source
        1..325
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3432289"
            /clone_lib="NCI-CGAP_Pr28"
            /sex="male"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
            with a modified polylinker; Plasmid DNA from the
            normalized library NCI-CGAP_Pr22 was prepared, and ss
            circles were made in vitro. Following HAP purification,
            this DNA was used as tracer in a subtractive hybridization
            reaction. The driver was PCR-amplified cDNAs from a pool
            of 5,000 clones made from the same library (clonoids
            985608-986759, 1101192-1101959, and 1217928-1220615).
            Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 65 a 108 c 71 g 81 t
ORIGIN
    Query Match 81.0%; Score 17; DB 12; Length 325;
    Best Local Similarity 100.0%; Pred. No. 8.7e+02;
    Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 GCACCATGTGATCCCG 19
Db 147 GCACCATGTGATCCCG 131
    |||||||
RESULT 8
AV005955/c
LOCUS AV005955 156 bp mRNA linear EST 25-AUG-1999
DEFINITION AV005955 Mus musculus C57BL/6J heart Mus musculus cDNA clone
1020003M09, mRNA sequence.
ACCESSION AV005955
VERSION AV005955.1 GI:4782805
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 156)
AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara
, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Naitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y., Sugahara
, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N.,
Watanabe, S., Yagame, M., Yamamura, F., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@r.riken.go.jp
Thermotabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524. (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

```





Email: dmlton@biohpc.harvard.edu  
 Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,  
 2000) Library was constructed by Catherine Lee DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Marie Scearce  
 (mscearce@mail.med.upenn.edu)  
 Seq primer: -400P from Gibco.  
 Location/Qualifiers

#### FEATURES

source  
 1..204  
 /organism="Mus musculus"  
 /strain="129/Sv x CD1"  
 /db\_xref="taxon:10090"  
 /clone\_lib="Kaeatner ng3 wt"  
 /dev\_stage="P.C. 14.5"  
 /lab\_host="E. coli-DH12S (GIBCO)"  
 /note="Organ: pancreas; Vector: pSPORT1 (GIBCO); Site 1:  
 Not I; Site 2: Sal I; The library was prepared by  
 Catherine S. Lee and has not been published. The pancreas  
 was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,  
 2000). The cDNA's were prepared with an oligo containing a  
 NotI site, and SalI linkers were added to the ends. The  
 inserts were cut with NotI before being cloned into the  
 NotI-SalI sites in the vectors. This is one of two  
 libraries, ng3 wt and ng3 -/- The wt library is in  
 pSPORT1, T7 promoter is 5'-. The wt library is in  
 pSPORT1, 40 c 32 g 64 t

#### BASE COUNT

ORIGIN 68 a 40 c 32 g 64 t

Query Match 80.0%; Score 16.8; DB 14; Length 204;  
 Best Local Similarity 90.0%; Pred. No. 9.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps - 0;

QY 2 GGCACCATGTGATCCCGAG 21

|||||

Db 43 GGCACCATGTGATCCCGAG 62

#### RESULT 13

BG791763/c  
 LOCUS 207 bp mRNA linear EST 30-MAY-2001  
 DEFINITION UTSH\_H15E9 Adult Mouse Cardiac Muscle Library Mus musculus  
 cDNA clone UTSH\_H15E9, mRNA sequence.

ACCESSION BG791763

VERSION BG791763.1 GI:14127333

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1 (bases 1 to 207)

#### AUTHORS

Gallardo, T.D., Schageman, J.J., Pertsemilidis, A., Garner, H.R.,  
 Williams, R.S., and Shohet, R.V.

UT Southwestern Medical Center, Adult Mouse Cardiac cDNA Library

Unpublished (2001)

Contact: Schageman JJ

Shohet/Garner Labs

University of Texas Southwestern Medical Center

6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA

Tel: 214 648 1674

Email: Jeff.Schageman@UTSouthwestern.edu

cDNA library constructed by UTSW as a component of the Program for  
 Genomic Applications (PGA) and the Reynolds Heart Disease  
 Prevention grants for use in cDNA microarray experiments. Sequence  
 Quality: Sequence ends were trimmed based on percentage of ambigu  
 us base calls or 'N's in windowed segments. Sequencing: First-pass  
 sequencing; ABI Prism 377 sequencer and analysis software.  
 Seq primer: M13/pUC Reverse.

#### FEATURES

Location/Qualifiers

1..207

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_lib="UTSW\_H15E9"

/clone\_lib="UTSW Adult Mouse Cardiac Muscle Library"

/sex="Pooled"

/tissue\_type="Cardiac muscle"

/dev\_stage="2 months"

/lab\_host="DH5a"

/note="Vector: pAMP10 (Gibco); Cloned unidirectionally.  
 Primer: Oligo dT. RNA Isolation: cytoplasmic RNA preps  
 (Manniat); Cloning Technique: CUA Cloning (CloneAmp,  
 Life Technologies); Average insert size: 1.8 Kb;  
 Insertion site: TAGTCCCATGATTCGAGT---. Other  
 information regarding entire library may be found at  
 http://pga.swmed.edu/Data/Libraries/microarray\_cdna\_librar  
 ies.htm."

BASE COUNT 88 a 31 c 37 g 48 t 3 others

#### ORIGIN

Query Match 80.0%; Score 16.8; DB 12; Length 207;

Best Local Similarity 90.0%; Pred. No. 9.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21

|||||

Db 93 GGCACCATGTGATCCCGAG 74

#### RESULT 14

BB185381/c

LOCUS 212 bp mRNA linear EST 30-JUN-2000

DEFINITION BB185381 RIKEN full-length enriched, adult male spinal cord Mus.

musculus cDNA clone A330023I12 3' similar to U51167 Mus musculus

isocitrate dehydrogenase mRNA, mRNA sequence.

ACCESSION BB185381

VERSION BB185381.1 GI:8845952

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 212)

AUTHORS Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,

Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.

, Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.

, Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata

, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.

, Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya

, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.

, Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino

, M., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S., Sasaki

, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Thermotabilization and thermoactivation of thermostable enzymes by

trehalose and its application for the synthesis of full length

cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998).

Itoh, M., Kitsuaki, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki

, Y., and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation

system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P., and Hayashizaki, Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.FEATURES  
source

## Location/Qualifiers

1. .212  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A330023112"  
 /clone\_lib="RIKEN full-length enriched, adult male spinal cord"  
 /sex="male"  
 /tissue\_type="spinal cord"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCCAGAGTAAATTAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 55 a 39 c 40 g 78 t

## ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 212;

Best Local Similarity 90.0%; Pred. No. 9.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21

|||||

DB 185 GGCACCATGTGATCCAGAG 166

## RESULT 15

BB149886/c

LOCUS

BB149886 RIKEN full-length enriched, 6 days neonate skin Mus  
 musculus cDNA clone A030001P06 3', similar to U51167 Mus musculus  
 isocitrate dehydrogenase mRNA, mRNA sequence.

ACCESSION BB149886

VERSION BB149886.1

KEYWORDS GI:8904823

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 225)  
 Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Hara, A., Hayatsu, N.,  
 Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,  
 Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,  
 Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,  
 Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,  
 Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y.,  
 Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,  
 Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T.,  
 Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yamanaka, I.,  
 Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M.,  
 Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

COMMENT

Unpublished (2000)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

Email: [genome-res@gsc.riken.go.jp](mailto:genome-res@gsc.riken.go.jp),  
[URL:http://genome.gsc.riken.go.jp/](mailto:URL:http://genome.gsc.riken.go.jp/)  
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
 Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.  
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES  
source

## Location/Qualifiers

1. .225  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="A030001P06"  
 /clone\_lib="RIKEN full-length enriched, 6 days neonate skin"  
 /tissue\_type="skin"  
 /dev\_stage="6 days neonate"  
 /lab\_host="DH10B"  
 /note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCCAGAGTAAATTAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 54 a 43 c 48 g 80 t

## ORIGIN

Query Match 80.0%; Score 16.8; DB 10; Length 225;

Best Local Similarity 90.0%; Pred. No. 9.6e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCACCATGTGATCCCGAG 21

|||||

DB 197 GGCACCATGTGATCCAGAG 178

Search completed: July 8, 2003, 09:21:34

Job time: 1068.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 242.107 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-11  
Perfect score: 22  
Sequence: 1 gcaagcactagcttcagtgtg 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl :

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pin.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	22	100.0	22	6	AX014711	Sequence
2	22	100.0	2765	6	AX014701	Sequence
3	21	85.6	174953	2	AC094732	Rattus no
4	18.4	83.6	179141	2	AC019091	AC019091 Homo sapi
5	18.4	83.6	191911	9	AC098965	Homo sapi
6	17.8	80.9	110000	2	AL683889_2	Continuation (3 of
7	17.8	80.9	110724	9	AL137222	Human DNA
8	17.8	80.9	152464	2	AC027421	AC027421 Homo sapi
9	17.8	80.9	152602	2	AC108898	Felis cat
10	17.8	80.9	152625	2	AC107322	Human DNA
11	17.8	80.9	157243	9	AL353622	Homo sapi
12	17.8	80.9	157986	2	AC025860	Homo sapi
13	17.8	80.9	164018	2	AC074242	Homo sapi
14	17.8	80.9	181718	2	AC130917	Rattus no
15	17.8	80.9	184864	9	AC013553	Homo sapi
16	17.8	80.9	192777	2	AC095271	Rattus no
17	17.8	80.9	198935	2	AC068573	Homo sapi
18	17.8	80.9	244280	2	AC125756	Rattus no
19	17.8	80.9	349050	1	AP003586	Nostoc sp
20	17.4	79.1	30631	9	AL355141	Human DNA
21	17.4	79.1	89198	2	AC021365	Homo sapi
22	17.4	79.1	116736	2	AC095745	Rattus no
23	17.4	79.1	130910	8	OSJN00132	Orzya sat
24	17.4	79.1	140711	2	AC117187	Mus muscu
25	17.4	79.1	145908	2	AC117959	Rattus no
26	17.4	79.1	175199	2	AC098270	Rattus no
27	17.4	79.1	185605	2	AC068440	Homo sapi
28	17.4	79.1	187704	2	AC122024	Mus muscu
29	17.4	79.1	189359	2	AC124697	Mus muscu
30	17.4	79.1	200395	2	AC145589	Rattus no
31	17.4	79.1	213524	2	AL390792	Homo sapi
32	17.4	79.1	223911	2	AC123931	Mus muscu
33	17.4	79.1	224852	2	AL805899	Mus muscu
34	17.2	78.2	722	9	HS4330110	Homo sapi
35	17.2	78.2	4439	9	AF074333	Homo sapi
36	17.2	78.2	11474	2	AC024110	Mus muscu
37	17.2	78.2	53735	2	AC101365	Mus muscu
38	17.2	78.2	64940	2	AC119266	Mus muscu
39	17.2	78.2	75892	9	AC114302	Homo sapi
40	17.2	78.2	112895	2	AC125689	Rattus no
41	17.2	78.2	117362	2	AC094277	Rattus no
42	17.2	78.2	143052	2	AC098214	Rattus no
43	17.2	78.2	144429	2	AC121685	Rattus no
44	17.2	78.2	153225	2	AC013581	Homo sapi
45	17.2	78.2	153538	9	AC093279	Homo sapi

ALIGNMENTS

RESULT 1  
AX014711  
LOCUS AX014711  
DEFINITION Sequence 11 from Patent WO9953077.  
ACCESSION AX014711  
VERSION AX014711.1 GI:10040984  
KEYWORDS synthetic construct.  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1 (bases 1 to 22)  
AUTHORS Schwartz J.C., Gros C., Ouimet, T., Rose, C., Bonhomme, M.C. and  
Facchinetti, P.  
TITLE Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy.

linear PAT 07-SEP-2000

JOURNAL Patent: WO 9953077-A 11 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source  
1. .22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
BASE COUNT 6 a 5 c 6 g 5 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAAGCACTAGCTTCAGTGTG 22  
|||||  
Db 1 GCAAAGCACTAGCTTCAGTGTG 22

RESULT 2  
AX014701 2765 bp DNA linear PAT 07-SEP-2000  
LOCUS  
DEFINITION Sequence 1 from Patent W09953077.  
ACCESSION AX014701  
VERSION AX014701.1 GI:10040975  
KEYWORDS  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 2765)  
AUTHORS Schwartz,J.C., Gros,C., Oulmet,T., Rose,C., Bonhomme,M.C. and  
Facchinetti,P.  
TITLE Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
JOURNAL Patent: WO 9953077-A 11 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
source  
1. .2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107. .2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSSVGVMMERADNCGRRRLGFVEGGLLVLLTLLMGAIVTL  
GVPTSGKQPLNSLLHVRHETVVKRVLRDSQSDICTPSCVIAARIQNM  
QSKPCDNFYQACGGLRHVVPETNSRYSFVDIRLDEVLKGLVLEDSVQHRRA  
VERAKTLRSCMQSVTEKRSDPELVNLDVGMGPVAMDKWMTGPKWELEROLAV  
LNSQENRVLIDLFVNDONSRRHVIYIDQPTLGMPSEVYFKEDSHRVEAVLQEM  
TSVATMLRDLNLPGETDLVOEMAQVHLHLETHLANATVPOEKRDVTLYHRMGLN  
LQRFGLKGNFTLFIQNLVSSVOVELLPNEEVVYIGIPLYENLEEIIDVPEATLQ  
YLVRVLDRIGLSQKFEARVDVNDLNMDESKKKAQKALNIREQIGPDY  
IKRAFSDKSVISSEITEKIRSVFVNDLNMDESKKKAQKALNIREQIGPDY  
ILEDDNNRLIDEEYSLLTFSDLYFENGLQNNKAQSLKLRKQVQNLWIIGAAVY  
NATYSPNNRLVFPAGILOPFPSKQDQNALNFGGICGWIGHEITHFDGDNFRDN  
GNMLDWSNESARHFQSQCMYQIYNSWELADGNVNGFTLGENIADNGVQRA  
YKAYLOWLAEGRDQRLPGLNLTAYQLFFINAYQWCGSYRPEAFIOISIKTDVHSPLK  
YRVGLSLQNLPGFSEAFHCPRGSPMPHNNCR1W"

BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAAAGCACTAGCTTCAGTGTG 22

Db 1 GCAAAGCACTAGCTTCAGTGTG 22

RESULT 3  
AC094732 174953 bp DNA linear HTG 20-DEC-2001  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-516, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
56 unordered pieces.  
AC094732  
AC094732.2 GI:17941511  
HTG: HTGS\_PHASE1.  
KEYWORDS Rattus norvegicus.  
SOURCE Rattus norvegicus  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 174953)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burket,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Loulseghe,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,J., Ren,Y., Rives,M., Rojas,A., Rojuben,I., Rolfe,M.,  
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,  
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs.R.  
Direct Submission  
2 (bases 1 to 174953)  
Unpublished  
Worley,K.C.  
Direct Submission  
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15624568.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
----- Project name: GBSG

```

Center clone name: CH230-516
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 152255 bases at least Q40
Consensus quality: 158448 bases at least Q30
Consensus quality: 164461 bases at least Q20
Estimated insert size: 155965; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; aarose-fp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length.
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs, the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*

```

1	13782:	contig of 13782 bp in length
*	13783:	gap of unknown length
*	13882:	contig of 9405 bp in length
*	23887:	contig of 9405 bp in length
*	23887:	gap of unknown length
*	23888:	contig of 4694 bp in length
*	23888:	contig of 4694 bp in length
*	28181:	gap of unknown length
*	33907:	contig of 5626 bp in length
*	33907:	gap of unknown length
*	39271:	contig of 5364 bp in length
*	39271:	gap of unknown length
*	39372:	contig of 4899 bp in length
*	39372:	gap of unknown length
*	44370:	contig of 3353 bp in length
*	44371:	contig of 3353 bp in length
*	47724:	gap of unknown length
*	47724:	contig of 5604 bp in length
*	47824:	gap of unknown length
*	53428:	contig of 3776 bp in length
*	53428:	gap of unknown length
*	57304:	contig of 4615 bp in length
*	57404:	contig of 4615 bp in length
*	62019:	gap of unknown length
*	62119:	contig of 4558 bp in length
*	66677:	gap of unknown length
*	66677:	contig of 3425 bp in length
*	70201:	gap of unknown length
*	70201:	contig of 3781 bp in length
*	74082:	gap of unknown length
*	74182:	contig of 4124 bp in length
*	74183:	gap of unknown length
*	78307:	contig of 2816 bp in length
*	78407:	gap of unknown length
*	81223:	contig of 3230 bp in length
*	81223:	gap of unknown length
*	84552:	contig of 4187 bp in length
*	84552:	gap of unknown length
*	84553:	contig of 4187 bp in length
*	84553:	gap of unknown length
*	88940:	contig of 3317 bp in length
*	88940:	gap of unknown length
*	92257:	contig of 2639 bp in length
*	92257:	gap of unknown length
*	94995:	contig of 3370 bp in length
*	94995:	gap of unknown length
*	95096:	contig of 3067 bp in length
*	95096:	gap of unknown length
*	98466:	contig of 3207 bp in length
*	98466:	gap of unknown length
*	98566:	contig of 4209 bp in length
*	98566:	gap of unknown length
*	101633:	contig of 3087 bp in length
*	101633:	gap of unknown length
*	101733:	contig of 3087 bp in length
*	101733:	gap of unknown length
*	104940:	contig of 4209 bp in length
*	104940:	gap of unknown length
*	105040:	contig of 3087 bp in length
*	105040:	gap of unknown length
*	109249:	contig of 2141 bp in length
*	109349:	contig of 2141 bp in length
*	112436:	contig of 2141 bp in length
*	112436:	gap of unknown length
*	112536:	contig of 2791 bp in length
*	112536:	gap of unknown length
*	114677:	contig of 2791 bp in length
*	114677:	gap of unknown length
*	117568:	contig of 2791 bp in length
*	117568:	gap of unknown length

*	117668	119209: contig of 1542 bp in length
*	119210	119309: gap of unknown length
*	119310	121797: contig of 2488 bp in length
*	121798	121897: gap of unknown length
*	121898	124293: contig of 2396 bp in length
*	124294	124393: gap of unknown length
*	124394	126428: contig of 2035 bp in length
*	126429	126528: gap of unknown length
*	126529	129525: contig of 2997 bp in length
*	129526	129625: gap of unknown length
*	129626	132334: contig of 2709 bp in length
*	132335	132434: gap of unknown length
*	132435	135274: contig of 2840 bp in length
*	135275	135374: gap of unknown length
*	135376	138074: contig of 2700 bp in length
*	138075	138174: gap of unknown length
*	138175	139985: contig of 1811 bp in length
*	139986	140085: gap of unknown length
*	140087	142273: contig of 2188 bp in length
*	142274	142373: gap of unknown length
*	142375	143598: contig of 1225 bp in length
*	143599	143698: gap of unknown length
*	143699	145334: contig of 1736 bp in length
*	145340	145534: gap of unknown length
*	145535	146985: contig of 1451 bp in length
*	146986	147085: gap of unknown length
*	147086	148099: contig of 1014 bp in length
*	148100	148199: gap of unknown length
*	148200	150915: contig of 2716 bp in length
*	150916	151015: gap of unknown length
*	151017	152501: contig of 1486 bp in length
*	152502	152601: gap of unknown length
*	152603	154010: contig of 1409 bp in length
*	154011	154110: gap of unknown length
*	154111	155758: contig of 1048 bp in length
*	155759	155858: gap of unknown length
*	155859	157622: contig of 1764 bp in length
*	157630	157722: gap of unknown length
*	157723	159428: contig of 1706 bp in length
*	159429	159528: gap of unknown length
*	159529	161209: contig of 1681 bp in length
*	161210	161309: gap of unknown length
*	161310	163413: contig of 2104 bp in length
*	163414	163513: gap of unknown length
*	163514	164702: contig of 1189 bp in length
*	164703	164802: gap of unknown length
*	164803	165998: contig of 1196 bp in length
*	165999	166098: gap of unknown length
*	166099	167412: contig of 1314 bp in length
*	167413	167512: gap of unknown length
*	167513	169231: contig of 1719 bp in length
*	169232	169331: gap of unknown length
*	169332	170534: contig of 1203 bp in length
*	170535	170634: gap of unknown length
*	170635	172047: contig of 1413 bp in length
*	172048	172147: gap of unknown length
*	172148	173509: contig of 1362 bp in length
*	173510	173609: gap of unknown length
*	173610	174953: contig of 1344 bp in length.

## FEATURES

```
Query Match          95.5%; Score 21; DB 2; Length 174953;
Best Local Similarity 100.0%; Pred. No. 3.1;
```

Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CAAAGCACTAGCTTCAGTGTG 22  
 |||||  
 Db 83065 CAAAGCACTAGCTTCAGTGTG 83085

RESULT 4  
AC019091  
LOCUS

AC019091 . 179141 bp DNA linear HTG 16-JUL-2000

DEFINITION Homo sapiens chromosome 16 clone RP11-378B23, WORKING DRAFT  
 SEQUENCE, 35 unordered pieces.  
 ACCESSION AC019091  
 VERSION AC019091.3 GI:9230842  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 179141)  
 Waterston,R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 179141)  
 Waterston,R.H.  
 Direct Submission  
 Submitted (30-DEC-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Jul 16, 2000 this sequence version replaced gi:7023910.

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

----- Project Information -----

Center project name: H.NH0378B23

----- Summary Statistics -----

Sequencing vector: M13; 76%

Sequencing vector: plasmid; 24%

Chemistry: Dye-primer ET; 76% of reads

Chemistry: Dye-terminator Big Dye; 24% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 158833 bases at least Q40

Consensus quality: 165730 bases at least Q30

Consensus quality: 169500 bases at least Q20

Insert size: 191000; agarose-fp

Insert size: 175741; sum-of-contigs

Quality coverage: 3.34 in Q20 bases; agarose-fp

Quality coverage: 3.41 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1710: contig of 1710 bp in length  
 \* 1711 1810: gap of unknown length  
 \* 1811 2901: contig of 1091 bp in length  
 \* 2902 3001: gap of unknown length  
 \* 3002 4208: contig of 1207 bp in length  
 \* 4209 4308: gap of unknown length  
 \* 4309 6465: contig of 2157 bp in length  
 \* 6466 6565: gap of unknown length  
 \* 6566 7944: contig of 1379 bp in length  
 \* 7945 8044: gap of unknown length  
 \* 8045 9970: contig of 1926 bp in length  
 \* 9971 10071: gap of unknown length  
 \* 10072 11727: contig of 1656 bp in length  
 \* 11728 11826: gap of unknown length  
 \* 11827 13836: contig of 2010 bp in length  
 \* 13837 13936: gap of unknown length  
 \* 13937 15607: contig of 1671 bp in length  
 \* 15608 15707: gap of unknown length  
 \* 15708 17161: contig of 1454 bp in length  
 \* 17162 17261: gap of unknown length  
 \* 17262 19602: contig of 2341 bp in length  
 \* 19603 19702: gap of unknown length  
 \* 19703 22183: contig of 2481 bp in length  
 \* 22184 22283: gap of unknown length

22284 25598: contig of 3315 bp in length  
 \* 25599 25698: gap of unknown length  
 \* 25699 28926: contig of 3228 bp in length  
 \* 28927 29026: gap of unknown length  
 \* 29027 31649: contig of 2623 bp in length  
 \* 31650 31749: gap of unknown length  
 \* 31750 36421: contig of 4672 bp in length  
 \* 36422 36521: gap of unknown length  
 \* 36522 41712: contig of 5191 bp in length  
 \* 41713 41812: gap of unknown length  
 \* 41813 46441: contig of 4629 bp in length  
 \* 46442 46541: gap of unknown length  
 \* 46542 50218: contig of 3677 bp in length  
 \* 50219 50318: gap of unknown length  
 \* 50319 54292: contig of 3774 bp in length  
 \* 54293 54392: gap of unknown length  
 \* 54393 60118: contig of 5726 bp in length  
 \* 60119 60218: gap of unknown length  
 \* 60219 64595: contig of 4377 bp in length  
 \* 64596 64695: gap of unknown length  
 \* 64696 70204: contig of 5509 bp in length  
 \* 70205 70304: gap of unknown length  
 \* 70305 77397: contig of 7093 bp in length  
 \* 77398 77497: gap of unknown length  
 \* 77498 83379: contig of 5882 bp in length  
 \* 83380 83479: gap of unknown length  
 \* 83480 88925: contig of 5446 bp in length  
 \* 88926 89025: gap of unknown length  
 \* 89026 95730: contig of 6605 bp in length  
 \* 95731 95731: gap of unknown length  
 \* 95732 100873: contig of 5143 bp in length  
 \* 100874 100973: gap of unknown length  
 \* 100974 107582: contig of 6609 bp in length  
 \* 107583 107682: gap of unknown length  
 \* 107683 115744: contig of 8062 bp in length  
 \* 115744 115844: gap of unknown length  
 \* 115845 124829: contig of 8985 bp in length  
 \* 124830 124929: gap of unknown length  
 \* 124930 134099: contig of 9170 bp in length  
 \* 134100 134199: gap of unknown length  
 \* 134200 145093: contig of 10894 bp in length  
 \* 145094 145193: gap of unknown length  
 \* 145194 158694: contig of 13501 bp in length  
 \* 158695 158794: gap of unknown length  
 \* 158795 179141: contig of 20347 bp in length.

#### FEATURES

##### source

1. 179141  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-378B23"

BASE COUNT 51098 a 37432 c 36902 g 50295 t 3414 others  
 ORIGIN

##### Query Match

Best Local Similarity 83.6%; Score 18.4; DB 2; Length 179141;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAAAGCACTAGCTTCAGTGT 21

Db 167731 CAAAGCACTAGCTTCAGTGT 167750  
 ||||| ||||| ||||| |||||

##### RESULT 5

##### AC098965

##### LOCUS

DEFINITION Homo sapiens chromosome 16 clone RP11-419L9, complete sequence.

AC098965

AC098965

AC098965.2 GI:18483431

HTG.

KEYWORDS Homo sapiens.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

191911 bp DNA linear

PRI 04-FEB-2002

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 191911)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 191911)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (07-NOV-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 191911)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (04-FEB-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 4, 2002 this sequence version replaced gi:16756238.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

## FEATURES

source  
1. 191911  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-419L9"  
BASE COUNT 57916 a 41899 c 39920 g 52176 t  
ORIGIN  
Query Match 83.6%; Score 18.4; DB 9; Length 191911;  
Best Local Similarity 95.0%; Pred. No. 69;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAAAGCACTAGCTTCAGTGT 21  
||||||| |||||||

Db 85292 CAAAGCACTAGCTTCAGTGT 85311

RESULT 6  
AL683889 2/c  
WPCOMMENT  
Sequence split into 6 fragments LOCUS AL683889 Accession AL683889  
Fragment Name Begin End  
AL683889\_0 1 110000  
AL683889\_1 100001 210000  
AL683889\_2 200001 310000  
AL683889\_3 300001 410000  
AL683889\_4 400001 510000  
AL683889\_5 500001 562802  
Continuation (3 of 6) of AL683889 from base 200001 (AL683889 Homo sapiens chromosome 1 c

Query Match 80.9%; Score 17.8; DB:2; Length 110000;  
Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 CAAAGCACTAGCTTCAGTGT 22  
||||| ||||| |||||||

Db 85668 CAAAGCACTAGCTTCAGTGTG 85648

## RESULT 7

AL137222/c  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP1-209B5 on chromosome 6, complete  
sequence.  
ACCESSION  
AL137222  
VERSION  
AL137222.17 GI:11875997  
KEYWORDS  
HTG.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 110724)  
Direct Submission  
Submitted (15-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Dec 17, 2000 this sequence version replaced gi:11610952.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence has been finished according to sequence map criteria  
as follows. An attempt is made to resolve all sequencing problems,  
such as compressions and repeats, but not necessarily within known  
annotated repeat sequence elements. Where the sequence is  
ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C/elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP1-209B5 is from the library RP1-1 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone  
RP1-209B5. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RP1-209B5 is at 1 in this sequence. The  
true left end of clone RP3-479G13 is at 110626 in this sequence.  
The true right end of clone RP11-692C10 is at 90615 in this  
sequence.

## FEATURES

Location/Qualifiers  
1. 110724  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP1-209B5"  
/clone\_lib="RPC1-1"  
1. 464  
repeat\_region  
/note="L1M4 repeat: matches 3276. 3796 of consensus"  
475. 698  
repeat\_region  
/note="AluJo repeat: matches 85. 310 of consensus"  
705. 798  
repeat\_region  
/note="L1M4 repeat: matches 3176. 3266 of consensus"  
799. 1090  
repeat\_region  
/note="AluJo repeat: matches 1. 290 of consensus"  
1100. 1403  
repeat\_region  
/note="AluSp repeat: matches 1. 309 of consensus"  
1404. 1558  
repeat\_region  
/note="L1M4 repeat: matches 3000. 3169 of consensus"  
1644. 2340

```
repeat_region /note="L1MB3A repeat: matches 4721. .5420 of consensus"  
2341. .2605  
repeat_region /note="L1BP1 repeat: matches 5882. .6155 of consensus"  
2606. .3362  
repeat_region /note="L1MB3A repeat: matches 5420. .6150 of consensus"  
3363. .3542  
repeat_region /note="Alusq/x repeat: matches 126. .305 of consensus"  
3564. .3876  
repeat_region /note="Alusq repeat: matches 1. .312 of consensus"  
4210. .4527  
repeat_region /note="Alub repeat: matches 1. .312 of consensus"  
4896. .5221  
repeat_region /note="Aludo repeat: matches 1. .310 of consensus"  
6075. .6383  
repeat_region /note="Alusq repeat: matches 1. .311 of consensus"  
7034. .7346  
repeat_region /note="Alusx repeat: matches 1. .293 of consensus"  
8910. .9038  
repeat_region /note="FLAM_A repeat: matches 4. .132 of consensus"  
9257. .10254  
repeat_region /note="L1LP13 repeat: matches 634. .1651 of consensus"  
10346. .16471  
repeat_region /note="L1PA5 repeat: matches 1. .6143 of consensus"  
16481. .16600  
repeat_region /note="L1LMC5 repeat: matches 7786. .7911 of consensus"  
16707. .16841  
repeat_region /note="L1LP13 repeat: matches 1696. .1826 of consensus"  
16842. .17152  
repeat_region /note="Alusx repeat: matches 1. .312 of consensus"  
17153. .18399  
repeat_region /note="L1LP13 repeat: matches 1826. .3042 of consensus"  
18400. .18710  
repeat_region /note="Alusx repeat: matches 1. .312 of consensus"  
18711. .21713  
repeat_region /note="L1LP13 repeat: matches 3042. .6152 of consensus"  
21741. .21896  
repeat_region /note="MIR repeat: matches 12. .166 of consensus"  
21933. .22470  
repeat_region /note="L2 repeat: matches 1898. .2456 of consensus"  
22471. .22782  
repeat_region /note="Alusx repeat: matches 1. .312 of consensus"  
22783. .22967  
repeat_region /note="L2 repeat: matches 2456. .2634 of consensus"  
23148. .24767  
repeat_region /note="L1PA7 repeat: matches 4495. .6156 of consensus"  
24977. .25032  
repeat_region /note="28 copies 2 mer ca 96% conserved"  
25603. .25912  
repeat_region /note="Aluy repeat: matches 1. .310 of consensus"  
26684. .26728  
repeat_region /note="L1MB4 repeat: matches 6107. .6184 of consensus"  
26729. .26775  
repeat_region /note="MAD1 repeat: matches 34. .80 of consensus"  
26776. .27185  
repeat_region /note="L1MB4 repeat: matches 5676. .6107 of consensus"  
27411. .27541  
repeat_region /note="MER5B repeat: matches 48. .178 of consensus"  
27893. .28270  
repeat_region /note="MLT1A2 repeat: matches 1. .374 of consensus"  
28337. .28362  
repeat_region /note="13 copies 2 mer tg 100% conserved"  
28364. .28642  
repeat_region /note="Aludo repeat: matches 1. .280 of consensus"  
28651. .29320  
repeat_region /note="MER21B repeat: matches 4. .682 of consensus"  
29629. .29751  
repeat_region /note="Aluj repeat: matches 1. .123 of consensus"  
29787. .29985  
repeat_region /note="Aluj repeat: matches 118. .312 of consensus"  
30696. .31000  
repeat_region /note="MER58B repeat: matches 30. .341 of consensus"  
31554. .31865  
repeat_region /note="Alusx repeat: matches 1. .311 of consensus"
```

```
repeat_region 31952. .32217  
/note="Alub repeat: matches 1. .266 of consensus"  
32220. .32372  
repeat_region /note="THEIA repeat: matches 1. .354 of consensus"  
32590. .32667  
repeat_region /note="39 copies 2 mer aa 78% conserved"  
32929. .33213  
repeat_region /note="MLT1A1 repeat: matches 4. .361 of consensus"  
33228. .33385  
repeat_region /note="THEIC repeat: matches 1. .371 of consensus"  
34486. .34651  
repeat_region /note="MER58A repeat: matches 3. .184 of consensus"  
34709. .35062  
repeat_region /note="MERVL repeat: matches 2605. .2953 of consensus"  
35358. .35540  
repeat_region /note="Aludo repeat: matches 116. .290 of consensus"  
36184. .36279  
repeat_region /note="MER63 repeat: matches 835. .928 of consensus"  
36280. .36512  
repeat_region /note="L1MB8 repeat: matches 5927. .6167 of consensus"  
36519. .36568  
repeat_region /note="L1MB4 repeat: matches 5467. .5515 of consensus"  
36569. .36787  
repeat_region /note="Alusx repeat: matches 5. .222 of consensus"  
36789. .36883  
repeat_region /note="L1MB4 repeat: matches 5238. .5336 of consensus"  
36886. .36908  
repeat_region /note="HSMAR2 repeat: matches 1276. .1299 of consensus"  
36909. .37205  
repeat_region /note="Aluy repeat: matches 1. .301 of consensus"  
37206. .37417  
repeat_region /note="HSMAR2 repeat: matches 1078. .1276 of consensus"  
37420. .37595  
repeat_region /note="Aluj repeat: matches 134. .309 of consensus"  
37596. .37893  
repeat_region /note="Alusx repeat: matches 10. .302 of consensus"  
37980. .38288  
repeat_region /note="Alusx repeat: matches 3. .303 of consensus"  
38289. .38410  
repeat_region /note="Aluj repeat: matches 5. .130 of consensus"  
38411. .38932  
repeat_region /note="HSMAR2 repeat: matches 3. .525 of consensus"  
38934. .39052  
repeat_region /note="L1MB4 repeat: matches 5151. .5239 of consensus"  
39053. .39342  
repeat_region /note="Alusq repeat: matches 1. .290 of consensus"  
39343. .39446  
repeat_region /note="L1MB4 repeat: matches 5038. .5151 of consensus"  
39447. .39607  
repeat_region /note="MER63 repeat: matches 902. .1062 of consensus"  
40751. .40919  
repeat_region /note="MER53 repeat: matches 3. .189 of consensus"  
41181. .41465  
repeat_region /note="L1MEC repeat: matches 1133. .1434 of consensus"  
41466. .41719  
repeat_region /note="MER44A repeat: matches 1. .258 of consensus"  
41720. .42019  
repeat_region /note="Alusx repeat: matches 1. .300 of consensus"  
42020. .42103
```

Query Match 80.9%; Score 17.8; DB 9; Length 110724;  
Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CAAGCACTAGCTTCAGTGTG 22  
|||||

Db 94438 CAAGCACTAGCTTCGTGTTG 94418  
|||||

RESULT 8

AC027421/c

LOCUS

DEFINITION Homo sapiens chromosome 1 clone RP11-1C6 map 1, WORKING DRAFT

AC027421

152464 bp

DNA

linear

HTG 26-MAY-2000

SEQUENCE, 37 unordered pieces.  
AC027421  
AC027421.3 GI:8077017  
HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
Homo sapiens  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
1 (bases 1 to 152464)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,  
Collymore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferrelira,P., FitzHugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 25, 2000 this sequence version replaced gi:7656797.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7564  
Center clone name: L\_C\_6  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 135290 bases at least Q40  
Consensus quality: 139114 bases at least Q30  
Consensus quality: 145106 bases at least Q20  
Insert size: 147000; agarose-fp  
Quality coverage: 3.3 in Q20 bases; agarose-fp  
Quality coverage: 3.3 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 37 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1011: contig of 1011 bp in length  
\* 1012 1111: gap of 100 bp  
\* 1112 2156: contig of 1045 bp in length  
\* 2157 2256: gap of 100 bp  
\*  
\* 2257 2370: contig of 114 bp in length  
\* 2371 2470: gap of 100 bp  
\* 2471 4334: contig of 1864 bp in length  
\* 4335 4434: gap of 100 bp  
\* 4435 5783: contig of 1349 bp in length  
\* 5784 5883: gap of 100 bp  
\* 5884 7653: contig of 1770 bp in length  
\* 7654 7753: gap of 100 bp  
\* 7754 9718: contig of 1965 bp in length  
\* 9719 9818: gap of 100 bp  
\* 9819 12558: contig of 2740 bp in length  
\* 12559 12658: gap of 100 bp  
\* 12659 14948: contig of 2290 bp in length  
\* 14949 15048: gap of 100 bp  
\* 15049 17177: contig of 2129 bp in length  
\* 17178 17277: gap of 100 bp  
\* 17278 19726: contig of 2449 bp in length  
\* 19727 19836: gap of 100 bp  
\* 19837 22448: contig of 2622 bp in length  
\* 22449 22548: gap of 100 bp  
\* 22549 24938: contig of 2390 bp in length  
\* 24939 25038: gap of 100 bp  
\* 25039 26839: contig of 1801 bp in length  
\* 26840 26939: gap of 100 bp  
\* 26940 29806: contig of 2867 bp in length  
\* 29807 29906: gap of 100 bp  
\* 29907 32183: contig of 2277 bp in length  
\* 32184 32283: gap of 100 bp  
\* 32284 34927: contig of 2644 bp in length  
\* 34928 35027: gap of 100 bp  
\* 35028 38175: contig of 3148 bp in length  
\* 38176 38275: gap of 100 bp  
\* 38276 41731: contig of 3456 bp in length  
\* 41732 41831: gap of 100 bp  
\* 41832 44344: contig of 2513 bp in length  
\* 44345 44444: gap of 100 bp  
\* 44445 47199: contig of 2755 bp in length  
\* 47200 47299: gap of 100 bp  
\* 47300 50907: contig of 3608 bp in length  
\* 50908 51007: gap of 100 bp  
\* 51008 55098: contig of 4091 bp in length  
\* 55099 55198: gap of 100 bp  
\* 55199 59560: contig of 4362 bp in length  
\* 59561 59660: gap of 100 bp  
\* 59661 64094: contig of 4434 bp in length  
\* 64095 64194: gap of 100 bp  
\* 64195 69630: contig of 5436 bp in length  
\* 69631 69730: gap of 100 bp  
\* 69731 75725: contig of 5995 bp in length  
\* 75726 75825: gap of 100 bp  
\* 75826 81160: contig of 5335 bp in length  
\* 81161 81260: gap of 100 bp  
\* 81261 85940: contig of 4680 bp in length  
\* 85941 86040: gap of 100 bp  
\* 86041 90738: contig of 4698 bp in length  
\* 90739 90838: gap of 100 bp  
\* 90839 96873: contig of 6035 bp in length  
\* 96874 96973: gap of 100 bp  
\* 96974 103333: contig of 6360 bp in length  
\* 10334 103433: gap of 100 bp  
\* 10344 111832: contig of 8399 bp in length  
\* 111833 111932: gap of 100 bp  
\* 111933 120415: contig of 8483 bp in length  
\* 120416 120515: gap of 100 bp  
\* 120516 130911: contig of 10396 bp in length  
\* 130912 131011: gap of 100 bp  
\* 131012 141212: contig of 10201 bp in length  
\* 141213 141312: gap of 100 bp  
\* 141313 152464: contig of 11152 bp in length.  
\*  
\* Location/Qualifiers  
\* 1. 152464  
\* /organism="Homo sapiens"  
\* /db\_xref="taxon:9606"  
\*  
FEATURES  
source

```

/misc_feature /chromosome="1"
/misc_feature /map="1"
/misc_feature /clone="RP11-106"
/misc_feature /clone_lib="RPC1-11 Human Male BAC"
/misc_feature 1..1011
/misc_feature 1112..2156
/misc_feature /note="assembly_fragment"
/misc_feature 2257..2370
/misc_feature /note="assembly_fragment"
/misc_feature clone_end:T7
/misc_feature vector_side:left"
/misc_feature 2471..4334
/misc_feature /note="assembly_fragment"
/misc_feature 4435..5783
/misc_feature /note="assembly_fragment"
/misc_feature 5884..7653
/misc_feature /note="assembly_fragment"
/misc_feature 7754..9718
/misc_feature /note="assembly_fragment"
/misc_feature 9819..12538
/misc_feature /note="assembly_fragment"
/misc_feature 12659..14948
/misc_feature /note="assembly_fragment"
/misc_feature 15049..17177
/misc_feature /note="assembly_fragment"
/misc_feature 17278..19726
/misc_feature /note="assembly_fragment"
/misc_feature 19827..22448
/misc_feature /note="assembly_fragment"
/misc_feature 22549..24938
/misc_feature /note="assembly_fragment"
/misc_feature 25039..26839
/misc_feature /note="assembly_fragment"
/misc_feature 26940..29806
/misc_feature /note="assembly_fragment"
/misc_feature 29907..32183
/misc_feature /note="assembly_fragment"
/misc_feature 32284..34927
/misc_feature /note="assembly_fragment"
/misc_feature 35028..38175
/misc_feature /note="assembly_fragment"
/misc_feature 38276..41731
/misc_feature /note="assembly_fragment"
/misc_feature 41832..44344
/misc_feature /note="assembly_fragment"
/misc_feature 44445..47199
/misc_feature /note="assembly_fragment"
/misc_feature 47300..50907
/misc_feature /note="assembly_fragment"
/misc_feature 51008..55098
/misc_feature /note="assembly_fragment"

```

```

Query Match 80.9%; Score 17.8; DB 2; Length 152464;
Best Local Similarity 90.5%; Pred.No.1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 CAAGACACTAGCTTCAGTGTG 22
      ||||| ||||| ||||| |||||
DB 126227 CAAGAACTAGATTGAGTGTG 126207

```

```

RESULT 9
AC108898
LOCUS AC108898 152602 bp DNA linear HTG 01-FEB-2002
DEFINITION Felis catus clone RP86-212B12, WORKING DRAFT SEQUENCE, 2 unordered
           pieces.
ACCESSION AC108898
VERSION AC108898.1 GI:18464074
KEYWORDS HTGS_PHASE1; HTGS_DRAFT.
SOURCE Felis catus.
ORGANISM Felis catus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

## REFERENCE

1 (bases 1 to 152602)  
 Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,  
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,  
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,  
 Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-O.,  
 Legaspi,R., Maduro,Q.L., Maduro,V.B., Masiello,C., Maskeri,B.,  
 Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,  
 Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,  
 Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,  
 Young,A., Zhang,L.-H. and Green,E.D.

## TITLE

JOURNAL

Unpublished

2 (bases 1 to 152602)

Green,E.D.

Direct Submission

Submitted (01-FEB-2002) NIH Intramural Sequencing Center, 8717

Grovermont Circle, Gaithersburg, MD 20877, USA

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: http://www.nisc.nih.gov

Contact: nisc\_mouse@nhgri.nih.gov

----- Project Information

Center project name: azg

Center clone name: 212B12

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 152288 bases at least Q40

Consensus quality: 152351 bases at least Q30

Consensus quality: 152412 bases at least Q20

Insert size: 125000; agarose-fp

Insert size: 152502; sum-of-contigs

Quality coverage: 12.78x in Q20 bases; agarose-fp

Quality coverage: 10.48x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 2 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 65456: contig of 65456 bp in length

\* 65457 65556: gap of unknown length

\* 65557 152602: contig of 87046 bp in length.

-----

Location/Qualifiers

1..152602

/organism="Felis catus"

/db\_xref="taxon:9685"

/clone="RP86-212B12"

/clone\_lib="RP86"

1..65456

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:right"

65557..152602

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right"

BASE COUNT 47056 a 30239 c 30430 g 44777 t 100 others

ORIGIN

Query Match 80.9%; Score 17.8; DB 2; Length 152602;

Best Local Similarity 90.5%; Pred.No.1.5e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

-----

QY 2 CAAGACACTAGCTTCAGTGTG 22

||||| ||||| ||||| |||||

DB 58330 CAAGCAATAGCTTCAGGTG 58350

```

RESULT 10
AC107322
LOCUS
DEFINITION AC107322 152625 bp DNA linear HTG:18-JAN-2002
Felis catus clone RP86-198D15, WORKING DRAFT SEQUENCE, 4 unordered
pieces.
ACCESSION AC107322
VERSION AC107322.1 GI:18201784
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Felis catus.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 152625)
Blakesley, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Ho, S.-L., Idol, J.R., Karlins, E., Laric, P., Lee-Lin, S.-O.,
Legaspi, R., Maduro, Q.L., Maduro, V.B., Masiello, C., Maskeri, B.,
Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Prasad, A.,
Stantripop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgueon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 152625)
Green, E.D.
Direct Submission
Submitted (18-JAN-2002) NFI Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA.
----- Genome Center
Center: NHI Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@hgrl.nih.gov
----- Project Information
Center project name: azh
Center clone name: 198D15
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151720 bases at least Q40
Consensus quality: 151988 bases at least Q30
Consensus quality: 152043 bases at least Q20
Insert size: 146000; agarose-fp
Quality coverage: 11.37x in Q20 bases; agarose-fp
Quality coverage: 10.90x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 2970: contig of 2970 bp in length
* 2971: gap of unknown length
* 3071: contig of 39590 bp in length
* 42661: gap of 39590 bp in length
* 42661: gap of unknown length
* 42761: contig of 50984 bp in length
* 93745: gap of unknown length
* 93845: contig of 58781 bp in length.
*
* Location/Qualifiers
* 1..152625
* /organism="Felis catus"
* /db_xref="taxon:9685"
* /clone="RP86-198D15"
* /clone.lib="RP86"
* 1..2970
misc_feature
1..2970

```

```

/note="assembly_fragment
clone_end:SP6
vector_side:right"
misc_feature
3071..42660
/note="assembly_fragment"
42761..93744
/note="assembly_fragment"
93845..152625
/note="assembly_fragment
clone_end:T7
vector_side:left"
BASE COUNT 46228 a 30256 c 30622 g 45219 t 300 others
ORIGIN
Query Match 80.9%; Score 17.8; DB 2; Length 152625;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CAAGGCACTAGCTTCAGGTG 22
||||| ||||||| |||
Db 133173 CAAGCAATAGCTTCAGGGTG 133193
RESULT 11
AL353622/c
LOCUS AL353622 157243 bp DNA linear PRI 30-APR-2002
DEFINITION Human DNA sequence from clone RP5-1092A3 on chromosome 1, complete
sequence.
ACCESSION AL353622
VERSION AL353622.33 GI:20386810
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157243)
Hall, R.
Direct Submission
Submitted (30-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 2, 2002 this sequence version replaced gi:18614033.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em:, EMBL; Sw:,
SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP5-1092A3 is from the library RPC1-5 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
Location/Qualifiers
1..157243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
FEATURES
source
1..157243

```

BASE COUNT 42538 a 35890 c 36187 g 42628 t  
 ORIGIN  
 /clone="RP5-1092A3"  
 /clone\_lib="RC1-5"

Query Match 80.9%; Score 17.8; DB 9; Length 157243;  
 Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
 Matches: 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CAAGACACTAGTTCAGTGTG 22  
 ||||| ||||| ||||| |||||  
 Db 96762 CAAGAACTAGTTCAGTGTG 96742

## RESULT 12.

AC025860/c  
 LOCUS 157986 bp DNA linear HTG 11-APR-2000  
 DEFINITION Homo sapiens chromosome 1 clone RP11-692J20 map 1, WORKING DRAFT  
 SEQUENCE, 46 unordered pieces.  
 AC025860  
 AC025860.2 GI:7534047  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 157986)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collamore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J.,  
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
 Vasiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 11, 2000 this sequence version replaced gi:7249383.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: L7542  
 Center clone name: 592\_J20  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 133649 bases at least Q40  
 Consensus quality: 144985 bases at least Q30

Consensus quality: 149889 bases at least Q20  
 Insert size: 188000; agarose-fp  
 Insert size: 153486; sum-of-contigs  
 Quality coverage: 2.6 in Q20 bases; agarose-fp  
 Quality coverage: 3.1 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 46 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1195: contig of 1195 bp in length  
 1196 1295: gap of 100 bp  
 1296 2678: contig of 1383 bp in length  
 2679 2778: gap of 100 bp  
 2779 3800: contig of 1022 bp in length  
 3801 3900: gap of 100 bp  
 3901 5245: contig of 1345 bp in length  
 5246 5345: gap of 100 bp  
 5346 7244: contig of 1899 bp in length  
 7245 7344: gap of 100 bp  
 7345 8760: contig of 1416 bp in length  
 8761 8860: gap of 100 bp  
 8861 10651: contig of 1791 bp in length  
 10652 10751: gap of 100 bp  
 10752 12671: contig of 1920 bp in length  
 12672 12771: gap of 100 bp  
 12772 14545: contig of 1774 bp in length  
 14546 14645: gap of 100 bp  
 14646 16422: contig of 1777 bp in length  
 16423 16522: gap of 100 bp  
 16523 18705: contig of 2183 bp in length  
 18706 18805: gap of 100 bp  
 18806 20968: contig of 2163 bp in length  
 20969 21068: gap of 100 bp  
 21069 22557: contig of 1489 bp in length  
 22558 22657: gap of 100 bp  
 22658 24984: contig of 2327 bp in length  
 24985 25084: gap of 100 bp  
 25085 27456: contig of 2372 bp in length  
 27457 27556: gap of 100 bp  
 27557 29684: contig of 2128 bp in length  
 29685 32391: contig of 2607 bp in length  
 32392 32491: gap of 100 bp  
 32492 35124: contig of 2633 bp in length  
 35125 35224: gap of 100 bp  
 35225 37272: contig of 2048 bp in length  
 37273 37372: gap of 100 bp  
 37373 39808: contig of 2436 bp in length  
 39809 39908: gap of 100 bp  
 39909 42359: contig of 2451 bp in length  
 42360 42459: gap of 100 bp  
 42460 44965: contig of 2506 bp in length  
 44966 45065: gap of 100 bp  
 45066 48724: contig of 3659 bp in length  
 48725 48824: gap of 100 bp  
 48825 50421: contig of 1597 bp in length  
 50422 50521: gap of 100 bp  
 50522 53449: contig of 2928 bp in length  
 53450 53549: gap of 100 bp  
 53550 56316: contig of 2767 bp in length  
 56317 56416: gap of 100 bp  
 56417 60843: contig of 4427 bp in length  
 60844 60943: gap of 100 bp  
 60944 65036: contig of 4093 bp in length  
 65037 65136: gap of 100 bp  
 65137 68208: contig of 3072 bp in length  
 68209 68308: gap of 100 bp  
 68309 72496: contig of 4188 bp in length

```

* 72497 72596: gap of 100 bp
* 72597 77903: contig of 5307 bp in length
* 77904 78003: gap of 100 bp
* 78004 81058: contig of 3053 bp in length
* 81057 81156: gap of 100 bp
* 81157 84136: contig of 2980 bp in length
* 84137 84236: gap of 100 bp
* 84237 89747: contig of 5511 bp in length
* 89748 89847: gap of 100 bp
* 89848 93510: contig of 3663 bp in length
* 93511 93610: gap of 100 bp
* 93611 97669: contig of 4059 bp in length
* 97670 97769: gap of 100 bp
* 97770 103021: contig of 5252 bp in length
* 103022 103121: gap of 100 bp
* 103122 107614: contig of 4493 bp in length
* 107615 107714: gap of 100 bp
* 107715 113485: contig of 5771 bp in length
* 113486 113585: gap of 100 bp
* 113586 118831: contig of 5246 bp in length
* 118832 118931: gap of 100 bp
* 118932 123029: contig of 4098 bp in length
* 123030 123129: gap of 100 bp
* 123130 127445: contig of 4316 bp in length
* 127446 127545: gap of 100 bp
* 127546 134672: contig of 7127 bp in length
* 134673 134772: gap of 100 bp
* 134773 142137: contig of 7365 bp in length
* 142138 142237: gap of 100 bp
* 142238 148711: contig of 6474 bp in length
* 148712 148811: gap of 100 bp
* 148812 157986: contig of 9175 bp in length.

```

## FEATURES

```

source
  1. 157986
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="1"
    /map="1"
    /clone="RP11-692J20"
    /clone_lib="RPC1-11 Human Male BAC"
  1. 1195
    /note="assembly_fragment"
    1296..2678
    /note="assembly_fragment"
    2779..3800
    /note="assembly_fragment"
    3901..5245
    /note="assembly_fragment"
    5346..7244
    /note="assembly_fragment"
    7345..8760
    /note="assembly_fragment"
    8861..10651
    /note="assembly_fragment"
    10752..12671
    /note="assembly_fragment"
    12772..14545
    /note="assembly_fragment"
    14646..16422
    /note="assembly_fragment"
    16523..18705
    /note="assembly_fragment"
    18806..20968
    /note="assembly_fragment"
    21069..22557
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right
    22658..24984
    /note="assembly_fragment"

```

Query Match 80.9%; Score 17.8; DB 2; Length 157986;  
 Best Local Similarity 90.5%; Pred. No. 1.5e+02;

```

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CAAGCACTAGCTTCAGTGTG 22
    ||||| ||||| ||||| |||||
DB 51743 CAAAGAAGTATGATTCAGTGTG 51723

RESULT 13
AC074242/c 164018 bp DNA linear HTG 07-AUG-2000
LOCUS Homo sapiens chromosome 6 clone RP11-112M16, WORKING DRAFT
DEFINITION AC074242
ACCESSION AC074242.2 GI:9665200
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 164018)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
Waterston,R.H.
Direct Submission
Submitted (22-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 3, 2000 this sequence version replaced gi:9369560.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0112M16
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 149547 bases at least Q40
Consensus quality: 153517 bases at least Q30
Consensus quality: 155530 bases at least Q20
Insert size: 152000; agarose-fp
Insert size: 160718; sum-of-contigs
Quality coverage: 4.14 in Q20 bases; agarose-fp
Quality coverage: 4.00 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1241: contig of 1241 bp in length
* 1242 1341: gap of unknown length
* 1342 2412: contig of 1071 bp in length
* 2413 2512: gap of unknown length
* 2513 3663: contig of 1151 bp in length
* 3664 3763: gap of unknown length
* 3764 4933: contig of 1170 bp in length
* 4934 5033: gap of unknown length
* 5034 6403: contig of 1369 bp in length
* 6403 7799: gap of unknown length
* 7799 7899: contig of 1297 bp in length
* 7899 9099: gap of unknown length
* 9099 9199: contig of 1200 bp in length
* 9199 9100: gap of unknown length

```

```

* 9200 10792: contig of 1593 bp in length
* 10793 10892: contig of unknown length
* 10893 13620: contig of 2728 bp in length
* 13621 13720: gap of unknown length
* 13721 15777: contig of 2057 bp in length
* 15778 15877: gap of unknown length
* 15878 18785: contig of 2908 bp in length
* 18786 18886: gap of unknown length
* 18886 21499: contig of 2614 bp in length
* 21500 21600: gap of unknown length
* 21600 23022: contig of 1423 bp in length
* 23023 23123: gap of unknown length
* 23123 25771: contig of 2649 bp in length
* 25772 25872: gap of unknown length
* 25872 28277: contig of 2406 bp in length
* 28278 28378: gap of unknown length
* 28378 31877: contig of 3500 bp in length
* 31878 31977: gap of unknown length
* 31978 36493: contig of 4516 bp in length
* 36494 36594: gap of unknown length
* 36594 39693: contig of 3099 bp in length
* 39693 39793: gap of unknown length
* 39793 43856: contig of 4064 bp in length
* 43857 43957: gap of unknown length
* 43957 47879: contig of 3923 bp in length
* 47880 51483: gap of unknown length
* 51484 51584: gap of unknown length
* 51584 56681: contig of 5098 bp in length
* 56682 56782: gap of unknown length
* 56782 62551: contig of 5770 bp in length
* 62552 62652: gap of unknown length
* 62652 67979: contig of 5328 bp in length
* 67980 68080: gap of unknown length
* 68080 74094: contig of 6015 bp in length
* 74095 74195: gap of unknown length
* 74195 80991: contig of 6797 bp in length
* 80992 81092: gap of unknown length
* 81092 87205: contig of 6114 bp in length
* 87206 87306: gap of unknown length
* 87306 92652: contig of 5347 bp in length
* 92653 92753: gap of unknown length
* 92753 103013: contig of 10260 bp in length
* 103013 103113: gap of unknown length
* 103113 110851: contig of 7739 bp in length
* 110852 110952: gap of unknown length
* 110952 120015: contig of 9064 bp in length
* 120016 120116: gap of unknown length
* 120116 130002: contig of 9887 bp in length
* 130003 130103: gap of unknown length
* 130103 145313: contig of 15211 bp in length
* 145314 145414: gap of unknown length
* 145414 164018: contig of 18605 bp in length.

```

## FEATURES

```

Source
1. 164018
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="6"
   /clone="RP11-112M16"

```

```

misc_feature
1. 1241
   /note="assembly_name:Contig28"
misc_feature
1342..2412
   /note="assembly_name:Contig29"
misc_feature
2513..3663
   /note="assembly_name:Contig31"
misc_feature
3764..4933
   /note="assembly_name:Contig32"
misc_feature
5034..6402
   /note="assembly_name:Contig33"
misc_feature
6503..7799
   /note="assembly_name:Contig34"
misc_feature
7900..9099
   /note="assembly_name:Contig36"

```

```

misc_feature
9200..10792
   /note="assembly_name:Contig37"
misc_feature
10893..13620
   /note="assembly_name:Contig38"
misc_feature
13721..15777
   /note="assembly_name:Contig40"
misc_feature
15878..18785
   /note="assembly_name:Contig41"
misc_feature
18886..21499
   /note="assembly_name:Contig42"
misc_feature
21600..23022
   /note="assembly_name:Contig43"
misc_feature
23123..25771
   /note="assembly_name:Contig44"
misc_feature
25872..28277
   /note="assembly_name:Contig45"
misc_feature
28378..31877
   /note="assembly_name:Contig46"
misc_feature
31978..36493
   /note="assembly_name:Contig47"
misc_feature
36594..39692
   /note="assembly_name:Contig48"
misc_feature
39793..43856
   /note="assembly_name:Contig49"
misc_feature
43957..47879
   /note="assembly_name:Contig50"
misc_feature
47980..51483
   /note="assembly_name:Contig51"
misc_feature
51584..56681
   /note="assembly_name:Contig52"
misc_feature
56782..62551
   /note="assembly_name:Contig53"
misc_feature
62652..67979
   /note="assembly_name:Contig54"
misc_feature
68080..74094
   /note="assembly_name:Contig55"
misc_feature
74195..80991
   /note="assembly_name:Contig56"
misc_feature
81092..87205
   /note="assembly_name:Contig57"
misc_feature
87306..92652
   /note="assembly_name:Contig58"
misc_feature
92753..103012
   /note="assembly_name:Contig59"
misc_feature
103113..110851
   /note="assembly_name:Contig60"
misc_feature
110952..120015
   /note="assembly_name:Contig61"
misc_feature
120116..130002
   /note="assembly_name:Contig62"
misc_feature
130103..145313
   /note="assembly_name:Contig63"
misc_feature
145414..164018
   /note="assembly_name:Contig64"

```

```

BASE COUNT 51062 a 31307 c 31421 g 46923 t 3305 others
ORIGIN

```

```

Query Match      80.9%; Score 17.8; DB 2; Length 164018;
Best Local Similarity 90.5%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 CAAAGCACTAGCTTCAGTGTG 22
   |||||
DB 27866 CAAAGCACTAGCTTCAGTGTG 27846

```

## RESULT 14

```

AC130917
LOCUS AC130917 181718 bp DNA linear HTG 15-AUG-2002
DEFINITION Rattus norvegicus clone CH230-237p19, *** SEQUENCING IN PROGRESS
ACCESSION AC130917
VERSION AC130917.1 GI:22261958

```

KEYWORDS  
SOURCE  
ORGANISM

HTG: HTGS\_PHASE1.

Norway rat.

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.REFERENCE  
AUTHORS

1 (bases 1 to 181718)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chaver, D., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Duigan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuheva, L., Loulsegged, H., Lozando, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathew, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okwuonu, G., Olarnpungsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quirroz, J., Rachlin, E., Reeves, K., Reigler, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 181718)

Rat Genome Sequencing Consortium.

Direct Submission

## JOURNAL

AUTHORS.

Submitted (15-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: KBFR

Center clone name: CH230-237p19

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 120233 bases at least Q40

Consensus quality: 130616 bases at least Q30

Consensus quality: 138732 bases at least Q20

\*\*\*\*\*

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 66 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1080: contig of 1080 bp in length  
1180: gap of unknown length  
1181: contig of 1027 bp in length  
2208: gap of unknown length  
2307: gap of unknown length  
3367: contig of 1060 bp in length  
3468: gap of unknown length  
4998: contig of 1531 bp in length  
5098: gap of unknown length  
6226: contig of 1128 bp in length  
6327: gap of unknown length  
6327: contig of 1017 bp in length  
7343: gap of unknown length  
7444: contig of 1136 bp in length  
8580: gap of unknown length  
8679: gap of unknown length  
8880: contig of 1255 bp in length  
9935: gap of unknown length  
10034: gap of unknown length  
11634: contig of 1600 bp in length  
11734: gap of unknown length  
12960: contig of 1226 bp in length  
13060: gap of unknown length  
14731: contig of 1671 bp in length  
14831: gap of unknown length  
16188: contig of 1357 bp in length  
16288: gap of unknown length  
16289: gap of unknown length  
18400: contig of 2112 bp in length  
18500: gap of unknown length  
19609: contig of 1109 bp in length  
19709: gap of unknown length  
20969: contig of 1260 bp in length  
21069: gap of unknown length  
22833: contig of 1764 bp in length  
22933: gap of unknown length  
24333: contig of 1400 bp in length  
24433: gap of unknown length  
25712: contig of 1279 bp in length  
25812: gap of unknown length  
27646: contig of 1834 bp in length  
27746: gap of unknown length  
29291: contig of 1545 bp in length  
29391: gap of unknown length  
31495: contig of 2104 bp in length  
31595: gap of unknown length  
33249: contig of 1654 bp in length  
33349: gap of unknown length  
35400: contig of 2051 bp in length  
35500: gap of unknown length  
36917: contig of 1417 bp in length  
37017: gap of unknown length  
38649: contig of 1632 bp in length  
38749: gap of unknown length  
41191: contig of 2442 bp in length  
41291: gap of unknown length  
43352: contig of 2061 bp in length  
43452: gap of unknown length  
46086: contig of 2634 bp in length  
46186: gap of unknown length  
47600: contig of 1414 bp in length  
46187



```
repeat_region /rpt_family="L2"
repeat_region 2215..2322
repeat_region /rpt_family="L2"
repeat_region 2673..2700
repeat_region /rpt_family="AT-rich"
repeat_region 2742..2856
repeat_region /rpt_family="MIR3"
repeat_region 2888..3180
repeat_region /rpt_family="AluSp"
repeat_region 3289..3597
repeat_region /rpt_family="AluY"
repeat_region 3936..4107
repeat_region /rpt_family="FAM"
repeat_region 4158..4204
repeat_region /rpt_family="L2"
repeat_region 4379..4492
repeat_region /rpt_family="MIR"
repeat_region complement(4500..4543)
repeat_region /rpt_family="MIR"
repeat_region complement(4544..4846)
repeat_region /rpt_family="AluSx"
repeat_region complement(4847..4972)
repeat_region /rpt_family="MIR"
repeat_region 5225..5333
repeat_region /rpt_family="MIR3"
repeat_region 5363..5655
repeat_region /rpt_family="LIMASx"
repeat_region complement(5656..5781)
repeat_region /rpt_family="MLT2D"
repeat_region 5790..6207
repeat_region /rpt_family="LIMASx"
repeat_region complement(6313..6695)
repeat_region /rpt_family="MLT2B1"
repeat_region complement(6696..6998)
repeat_region /rpt_family="AluJb"
repeat_region complement(6999..7108)
repeat_region /rpt_family="MLT2B1"
repeat_region complement(7117..7273)
repeat_region /rpt_family="MIR"
repeat_region complement(7485..7608)
repeat_region /rpt_family="MER5A"
repeat_region complement(8299..8609)
repeat_region /rpt_family="AluSp"
repeat_region complement(9109..9192)
repeat_region /rpt_family="MIR"
repeat_region complement(9302..9643)
repeat_region /rpt_family="LIPAI0"
repeat_region 9647..9916
repeat_region /rpt_family="L2"
repeat_region 10055..10271
repeat_region /rpt_family="L2"
repeat_region complement(10361..10480)
repeat_region /rpt_family="MER94"
repeat_region 10715..11025
repeat_region /rpt_family="AluJo"
repeat_region 11252..11563
repeat_region /rpt_family="AluY"
repeat_region 11818..12119
repeat_region /rpt_family="AluSg"
repeat_region complement(12564..12623)
repeat_region /rpt_family="MADEI"
repeat_region 12710..13002
repeat_region /rpt_family="AluSx"
repeat_region 12967..12977
repeat_region /note="<30 qual SNGI region"
repeat_region 13004..13008
repeat_region /note="<30 qual SNGI region"
repeat_region complement(13154..13452)
repeat_region /rpt_family="AluSp"
repeat_region complement(14745..15042)
repeat_region /rpt_family="AluSx"
repeat_region complement(15514..15810)
repeat_region /rpt_family="AluY"
```

```
repeat_region 16087..16465
repeat_region /rpt_family="MER61D"
repeat_region complement(16517..16762)
repeat_region /rpt_family="LIMC4a"
repeat_region complement(17245..17342)
repeat_region /rpt_family="LIMB1"
repeat_region complement(17599..17737)
repeat_region /rpt_family="MIR3"
repeat_region 18456..18775
repeat_region /rpt_family="L2"
repeat_region complement(18776..19087)
repeat_region /rpt_family="AluJb"
repeat_region 19088..19313
repeat_region /rpt_family="L2"
repeat_region complement(19342..20175)
repeat_region /rpt_family="LIMB7"
repeat_region 20176..20477
repeat_region /rpt_family="AluSx"
repeat_region complement(20478..20620)
repeat_region /rpt_family="LIMB7"
repeat_region complement(20621..20931)
repeat_region /rpt_family="AluJb"
repeat_region complement(20932..21437)
repeat_region /rpt_family="LIMB7"
repeat_region 21438..21459
repeat_region /rpt_family="(TAA)n"
repeat_region complement(21460..21545)
repeat_region /rpt_family="LIMB7"
repeat_region complement(21546..21834)
repeat_region /rpt_family="AluSx"
repeat_region complement(21835..22416)
repeat_region /rpt_family="LIMB7"
repeat_region complement(22417..22548)
repeat_region /rpt_family="FLAM_A"
repeat_region complement(22549..22796)
repeat_region /rpt_family="LIMB7"
repeat_region 22797..23107
repeat_region /rpt_family="AluSx"
repeat_region complement(23108..23247)
repeat_region /rpt_family="LIMB7"
repeat_region 23255..23457
repeat_region /rpt_family="L2"
```

Query Match 80.9%; Score 17.8; DB 9; Length 184864;  
Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAAGACCACTAGCTTCAGTGT 21  
Db 71781 GCACAGTACTAGCTTCAGTGT 71761

Search completed: July 8, 2003, 03:34:44  
Job time : 247.107 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic, - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 138.224 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-11

Perfect score: 22

Sequence: 1 gcaagcactagcttcagtgtg 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0 .

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	AAZ28817	Rat membrane metal
2	22	100.0	2765	AAZ28810	Rat membrane metal
3	16.8	76.4	331	ABK62554	Rat sequence diffe
4	16.8	76.4	2336	ABL25670	Drosophila melanog
5	16.8	76.4	7444	AAD32364	Human lung specifi
6	16.8	76.4	30620	AAK66931	Human immune/haema
7	16.4	74.5	563	AAV26013	Bacteriophage DNA
8	16.2	73.6	295	ABL73130	Corn tassell-derive
9	16.2	73.6	351	AAC30329	Human secreted pro

c 10	16.2	73.6	2022	21	AAA96492	cDNA encoding a hu
c 11	16.2	73.6	2226	24	ABL68244	Kidney cancer rela
c 12	16.2	73.6	2226	24	ABL68864	Kidney cancer rela
c 13	16.2	73.6	2274	24	ABN79910	Fungal ZRC gene se
c 14	16.2	73.6	2340	20	AAK84531	Pl artificial chro
c 15	16.2	73.6	2840	24	ABN59653	cDNA encoding neut
c 16	16.2	73.6	2925	21	AAA63763	EIF-2alpha kinase
c 17	16.2	73.6	3794	24	ABK24531	Human low adenosin
c 18	16.2	73.6	7207	21	AAF21316	Human adenosine re
c 19	16.2	73.6	7207	21	AAF21317	Human adenosine re
c 20	16.2	73.6	15630	21	AAF21317	Human low adenosin
c 21	16.2	73.6	15630	21	AAF21317	Human adenosine re
c 22	16.2	73.6	25701	22	AAK35195	Human reproductive
c 23	16.2	73.6	25758	22	AAK07077	Human reproductive
c 24	15.8	71.8	255	16	AAK25920	Human gene signatu
c 25	15.8	71.8	535	22	AAK22089	Human breast cance
c 26	15.8	71.8	536	22	AAK16019	Human breast cance
c 27	15.8	71.8	536	22	AAK16019	Human breast cance
c 28	15.8	71.8	1036	22	AAK24863	Human breast cance
c 29	15.8	71.8	1119	23	ABL08935	Drosophila melanog
c 30	15.8	71.8	1122	18	AAK05129	Sequence encoding
c 31	15.8	71.8	1122	18	AAK62690	Lactate oxidase va
c 32	15.8	71.8	1122	18	AAK62691	Lactate oxidase va
c 33	15.8	71.8	1122	18	AAK62692	Lactate oxidase va
c 34	15.8	71.8	1825	20	AAK20231	Aerococcus lactate
c 35	15.8	71.8	1825	24	ABN98216	E faecal is EPI24 g
c 36	15.8	71.8	1947	22	ABA5830	Human breast cell
c 37	15.8	71.8	1947	22	ABA56345	Human foetal liver
c 38	15.8	71.8	1947	22	ABA25986	Probe #4452 for ge
c 39	15.8	71.8	1947	22	AAK04524	Human brain expres
c 40	15.8	71.8	1947	22	AAK30025	Human bone marrow
c 41	15.8	71.8	1947	22	AAI14616	Probe #4549 for ge
c 42	15.8	71.8	1947	22	AAI35988	Probe #4674 used t
c 43	15.8	71.8	1947	22	AAI04430	Probe #4421 used t
c 44	15.8	71.8	1947	24	ABSO4597	Human genome-deriv
c 45	15.8	71.8	2301	24	ABL40503	Human serine/threo

#### ALIGNMENTS

RESULT 1  
AAZ28817  
ID AAZ28817 standard; DNA; 22 BP.  
XX  
AC AAZ28817;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Rat membrane metalloprotease NEPII gene probe #7.  
XX  
DE Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
XX  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
XX  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
XX  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
XX  
OS Synthetic.  
XX  
OS Rattus rattus.  
XX  
PN FR2777291-A1.  
XX  
PD 15-OCT-1999.  
XX  
PF 08-APR-1998; 98FR-0004389.  
XX  
PR 08-APR-1998; 98FR-0004389.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
XX  
PI Schwartz JC;  
XX  
DR WPI.; 1999-593429/51.

XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease  
XX  
PS Claim 3; Page 21; 29pp; French.  
XX  
CC Sequences AAZ28811-228827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysine II (NEPII) gene  
CC (AAZ28810). NEPII is involved in (in)activation of neuronal and hormonal  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
CC disorders.  
XX  
XX Sequence 22 BP; 6 A; 5 C; 6 G; 5 T; 0 other;  
SQ  
Query Match 100.0%; Score 22; DB 20; Length 22;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCAAAGCAGCTAGCTTCAGTGTG 22  
Db 1 GCAAAGCAGCTAGCTTCAGTGTG 22  
|||||  
RESULT 2  
AAZ28810  
ID AAZ28810 standard; cDNA; 2765 BP.  
XX  
XX AAZ28810;  
XX  
XX 01-FEB-2000 (first entry)  
DT  
XX  
XX Rat membrane metalloprotease NEPII gene.  
DE  
XX  
XX Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
XX  
XX Rattus rattus.  
OS  
XX  
XX FR2777291-AL.  
PN  
XX  
XX 15-OCT-1999.  
PD  
XX  
XX 08-APR-1998; 98FR-0004389.  
PF  
XX  
XX 08-APR-1998; 98FR-0004389.  
PR  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA  
XX  
XX Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
PI  
XX  
XX WPI; 1999-593429/51.  
DR  
XX  
XX P-PSDB; AAY44177.  
DR  
XX  
XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease  
XX  
XX Claim 2; Page 12-16; 29pp; French.  
PS  
XX  
CC This sequence represents the gene for the rat membrane metalloprotease  
CC designated neprilysine II (NEPII), which is involved in (in)activation  
CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
CC for specific substrates (used to detect NEPII in cells and tissues) or  
CC inhibitors, which can also be used to detect NEPII or for treatment of

CC disorders related to peptidergic signalling in which NEPII is involved,  
CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
CC endocrine disorders.  
XX  
XX Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
SQ  
Query Match 100.0%; Score 22; DB 20; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 0.45;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCAAAGCAGCTAGCTTCAGTGTG 22  
Db 1 GCAAAGCAGCTAGCTTCAGTGTG 22  
|||||  
RESULT 3  
ABK62554  
ID ABK62554 standard; cDNA; 331 BP.  
XX  
XX ABK62554;  
XX  
XX 18-JUN-2002 (first entry)  
DT  
XX  
XX Rat sequence differentially expressed in response to a hepatotoxin #461.  
DE  
XX  
XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
KW differential expression; centrilobular necrosis; steatosis.  
KW  
XX  
XX Rattus norvegicus.  
OS  
XX  
XX WO200210453-A2.  
PN  
XX  
XX 07-FEB-2002.  
PD  
XX  
XX 30-JUL-2001; 2001WO-US23872.  
PF  
XX  
XX 31-JUL-2000; 2000US-222040P.  
PR  
XX  
XX 02-NOV-2000; 2000US-244880P.  
PR  
XX  
XX 11-MAY-2001; 2001US-290029P.  
PR  
XX  
XX 15-MAY-2001; 2001US-290645P.  
PR  
XX  
XX 22-MAY-2001; 2001US-292336P.  
PR  
XX  
XX 06-JUN-2001; 2001US-29798P.  
PR  
XX  
XX 13-JUN-2001; 2001US-297457P.  
PR  
XX  
XX 19-JUN-2001; 2001US-298884P.  
PR  
XX  
XX 09-JUL-2001; 2001US-303459P.  
PR  
XX  
XX (GENE-) GENE LOGIC INC.  
PA  
XX  
XX Mendrick D, Porter MW, Johnson KR, Castile AL, Elashoff MR;  
PI  
XX  
XX WPI; 2002-241625/29.  
DR  
XX  
XX Predicting toxic effects of compounds or the progression of these toxic  
PT effects by determining the changes in gene expression in tissues or  
PT cells exposed to the toxin and comparing these to gene expression in  
PT unexposed tissues or cells  
XX  
XX Claim 1; Seq ID No 461; 239pp; English.  
PS  
XX  
XX The invention relates to methods for predicting toxic effects of  
CC compounds or the progression of these toxic effects by determining the  
CC global changes in gene expression in tissues or cells exposed to the  
CC toxin and comparing these to gene expression in unexposed tissues or  
CC cells. Also included are methods of predicting at least one toxic  
CC effect of a compound or progression of a toxic effect, preferably the  
CC hepatotoxicity of a compound, comprising detecting the level of  
CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression.  
CC The method can also be used to identify an agent which modulates the  
CC toxic response and predict cellular pathways that a compound modulates  
CC in a cell. The methods utilise a set of at least two probes (on a solid

CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridises to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in  
CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity  
CC is characterised by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent.  
XX  
SQ Sequence 331 BP; 95 A; 78 C; 70 G; 88 T; 0 other;

Query Match 76.4%; Score 16.8; DB 24; Length 331;  
Best Local Similarity 90.0%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGTG 22  
||||| || |||||  
Db 28 AAAGCAGCTAGCTTCAGTGTG 47

RESULT 4  
ID ABL25670 standard; DNA; 2336 BP.  
XX  
AC ABL25670;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 28483.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
PI WPT; 2001-656860/75.  
XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions  
XX  
PS Claim 1; SEQ ID NO 28483; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2336 BP; 638 A; 505 C; 529 G; 664 T; 0 other;

Query Match 76.4%; Score 16.8; DB 23; Length 2336;  
Best Local Similarity 90.0%; Pred. No. 1.6e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CAAAGCAGCTAGCTTCAGTGT 21  
||||| |||||  
Db 198 CAAAGCAGCTAGCTTCAGTGT 217

RESULT 5  
ID AAD32364 standard; DNA; 7444 BP.  
XX  
AC AAD32364;  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Human lung specific gene (LSG) #1.  
XX

XX Human; lung specific gene; LSG; lung embryonic development; cytostatic;  
KW lung cancer; vaccine; gene therapy; non-cancerous lung disease; gene;  
KW tumour; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200208278-A2.  
XX  
PD 31-JAN-2002.  
XX  
PF 20-JUL-2001; 2001WO-US22949.  
XX  
PR 21-JUL-2000; 2000US-219834P.  
XX  
PA (DIAD-) DIADEXUS INC.  
XX  
PI Macina RA, Nair M, Chen S;  
XX  
PI WPI; 2002-268964/31.  
XX

XX Novel lung specific gene useful for identifying, diagnosing,  
XX monitoring, staging, imaging and treating lung cancer and non-cancerous  
XX disease states in lung, for gene therapy, and for identifying lung  
XX tissue  
XX  
PS Claim 1; Page 150-153; 197pp; English.  
XX  
XX The present invention relates to lung specific genes (LSG) and their  
XX corresponding polypeptides. LSG is useful for identifying, diagnosing,  
XX monitoring, staging, imaging and treating lung cancer and non-cancerous  
XX disease states in lung, identifying lung tissue, monitoring and  
XX modifying lung embryonic development and differentiation, in gene  
XX therapy, as hybridisation probes, to detect LSG mRNA as a marker for  
XX lung cancer, as research reagents and materials for discovery of  
XX treatments and diagnostics to human disease, to detect complementary  
XX polynucleotides, and for chromosome identification. An antibody which  
XX binds LSG is useful to detect or image localisation of LSG in a patient  
XX for detecting or diagnosing a disease or condition, for preventing the  
XX onset and treatment of lung cancer, to isolate or to identify clones  
XX expressing LSG polypeptides, to purify LSG polypeptides, and to target  
XX tumours expressing LSG. The present sequence is human LSG DNA.  
XX  
SQ Sequence 7444 BP; 2172 A; 1679 C; 1695 G; 1898 T; 0 other;

Query Match 76.4%; Score 16.8; DB 24; Length 7444;  
Best Local Similarity 90.0%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAAGCACTAGCTTCACTG 20  
Db 6812 GCAAGCACTATCTACAGTG 6831  
RESULT 6  
AAK66931/c  
ID AAK66931 standard; DNA: 30620 BP.  
XX AAK66931;  
XX  
XX 06-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21743.  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01354.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225277.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 05-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0256719.  
 PR 08-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis.  
 XX  
 PS Disclosure; SEQ ID NO 21743; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent, the  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK34942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 30620 BP; 7407 A; 7216 C; 7192 G; 8805 T; 0 other;  
 Query Match 76.4%; Score 16.8; DB 22; Length 30620;  
 Best Local Similarity 90.0%; Pred. No. 2.3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GCAGAGCACTAGCTTCAGTG 20'  
 ||||| || ||||| |||||  
 Db 20724 GCAGAGCACTAGCTTCAGTG 20705  
 RESULT 7  
 AAV26013  
 ID AAV26013 standard; DNA; 563 BP.  
 XX  
 AC AAV26013;  
 XX  
 DT 28-AUG-1998 (first entry)  
 XX  
 DE Bacteriophage DNA SEQ ID NO:2-10.  
 XX  
 KW Bacteriophage; pathogen; Escherichia coli 0157; bactericide; food;  
 KW sterilisation; enteric haemorrhage; ds.  
 XX  
 OS Bacteriophage.  
 XX

PN WO9808944-A1.  
 XX  
 PD 05-MAR-1998.  
 XX  
 PF 26-AUG-1997; 97WO-JP02957.  
 XX  
 PR 19-APR-1997; 97JP-0135716.  
 PR 26-AUG-1996; 96JP-0261132.  
 PR 14-APR-1997; 97JP-0130236.  
 XX  
 PA (BIOV-) BIO VENTURE BANK CO LTD.  
 XX  
 PI Nishikori K, Takahashi S;  
 XX  
 DR WPI; 1998-230262/20.  
 XX  
 PT Bacteriophage with high specificity for particular pathogens such as  
 PT Escherichia coli 0157 - is incorporated in bactericides for food  
 PT sterilisation  
 XX  
 PS Claim 3; Page 39; 54pp; Japanese.  
 XX  
 CC The present sequence represents a DNA sequence from a novel  
 CC bacteriophage which has high specificity for pathogenic bacteria,  
 CC especially for strains of Escherichia coli causing enteric haemorrhage,  
 CC such as E. coli 0157. The bacteriophage can be incorporated into  
 CC bio-bactericidal compositions. These can contain more than one  
 CC bacteriophage strain in order to have a bactericidal effect against more  
 CC than one pathogen at the same time. The bio-bactericide may also contain  
 CC an amino-acid or other material to control the pH to 6.5-7.5 and ensure  
 CC stability of the bacteriophage, such as glycine, arginine or lysine.  
 CC The bio-bactericidal composition can be used for treating food (such as  
 CC meat or fish) to prevent bacterial contamination, e.g. in fresh food or  
 CC in the kitchens of restaurants, schools and other institutions. It may  
 CC be formulated as a spray. It can also be used to sterilise working  
 CC surfaces, aprons and other clothing. The bio-bactericides are safe to  
 CC humans but very potent against pathogenic bacteria.  
 XX  
 SQ Sequence 563 BP; 182 A; 93 C; 114 G; 169 T; 5 other;  
 Query Match 74.5%; Score 16.4; DB 19; Length 563;  
 Best Local Similarity 89.5%; Pred. No. 2.1e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 AAGCAGTACGCTTCAGTG 22  
 ||||| ||||| ||||| |||||  
 Db 328 AAGCAGTACGCTTCAGTG 346  
 RESULT 8  
 ABL73130  
 ID ABL73130 standard; cDNA; 295 BP.  
 XX  
 AC ABL73130;  
 XX  
 DT 14-MAY-2002 (first entry)  
 XX  
 DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:2504.  
 XX  
 KW Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPS;  
 KW inheritance; characteristic; growth; development; disease resistance;  
 KW environmental adaptability; quality; yield; molecular marker;  
 KW multigene trait; plant breeding; corn tassel; gene; ss.  
 XX  
 OS Zea mays.  
 XX  
 PN US2001051335-A1.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 16-APR-1999; 99US-0294093.  
 XX  
 PR 21-APR-1998; 98US-082567P.

XX (LALG/) LALGUDI R V.  
PA (ITOL/) ITO L Y.  
PA (SHER/) SHERMAN B K.  
XX Lalgudi RV, Ito LY, Sherman BK;  
XX WPI; 2002-163647/21.  
XX Novel purified corn tassel-derived polynucleotide useful for  
PT determining altered gene expression, to recover regulatory elements and  
PT to follow inheritance of desirable characteristics through hybrid  
PT breeding programs.  
XX  
PS Claim 1; SEQ ID 2504; 201pp; English.  
XX  
CC The present sequence describes a purified corn tassel-derived  
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)  
CC can be used for determining altered gene expression, to recover  
CC regulatory elements and to follow inheritance of desirable  
CC characteristics through hybrid breeding programs. (I) are also useful  
CC in the evaluation, and alteration of desired characteristics associated  
CC with growth and development, disease resistance, environmental  
CC adaptability, quality and yield, and as molecular markers for studying  
CC inheritance of multigene traits in a plant breeding program. (I) can be  
CC used to produce a tassel-specific profile of gene transcription, a  
CC transcript image, to clone regulatory elements for use in transformation  
CC vectors, to express a polypeptide, to identify, isolate or extend  
CC identical or related corn tassel nucleic acid sequences from DNA  
CC libraries, in nucleic acid hybridisation or amplification technologies,  
CC as query sequences to determine homology of known sequences, as probe  
CC for use in Southern or Northern hybridisation, and to identify the  
CC presence of and/or to determine the degree of similarity between two  
CC (or more) nucleic acid sequences.  
XX  
SQ Sequence 295 BP; 70 A; 63 C; 75 G; 86 T; 1 other;  
Query Match 73.6%; Score 16.2; DB 24; Length 295;  
Best Local Similarity 85.7%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 CAAGGCACTAGCTTCAGTGTG 22  
||||| ||| ||||| |||  
Db 93 CAAGTACTGCTTCAGAGTG 113  
  
RESULT 9  
AAC30329/c  
ID AAC30329 standard; cDNA; 351 BP.  
XX AAC30329;  
XX  
XX 06-OCT-2000 (first entry)  
XX Human secreted protein 5' EST, SEQ ID NO: 34404.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000EP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
DR

PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 34404; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors.  
XX  
SQ Sequence 351 BP; 96 A; 89 C; 90 G; 76 T; 0 other;  
Query Match 73.6%; Score 16.2; DB 21; Length 351;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 CAAGGCACTAGCTTCAGTGTG 22  
||||| ||| ||||| |||  
Db 251 CAAGGCACTATCATCAGTGTG 231  
  
RESULT 10  
AAA96492/c  
ID AAA96492 standard; cDNA; 2022 BP.  
XX AAA96492;  
XX  
XX 08-FEB-2001 (first entry)  
XX cDNA encoding a human transmembrane protein.  
XX  
KW Human; transmembrane protein; cell proliferation disorder; myeloma;  
KW reproductive disorder; smooth muscle disorder; neurological disorder;  
KW arteriosclerosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
KW allergy; ovulatory defect; angina; hypertension; stroke; epilepsy;  
KW Alzheimer's disease; Tourette's disorder; ss.  
XX  
OS Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 195..1319  
XX /\*tag= a  
XX  
XX WO200056891-A2.  
XX  
XX 28-SEP-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07817.  
XX  
XX 22-MAR-1999; 99US-0125537.  
XX  
XX 16-JUN-1999; 99US-0139565.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Yue H, Lal P, Tang YT, Hillman JL, Reddy R, Bandman O, Baughn MR;  
XX Lu DAM, Azimzai Y, Yang J;  
XX WPI; 2000-579485/54.  
XX  
DR

DR P-PSDB: AAB18979.

PT New human transmembrane proteins are used to treat a disease or

PT condition associated with decreased expression of functional HTMP e.g.

PT Tourette's disorder, angina and leukaemia

XX

PS Claim 4; Page 120-121; 130pp; English.

XX

CC The present sequence encodes a human transmembrane proteins (HTMP).

CC Agonists and antagonists of the protein are used to treat a disease

CC or condition associated with overexpression of the protein. Diseases

CC and conditions which can be treated include cell proliferative,

CC immunological, reproductive, smooth muscle and neurological disorders

CC e.g. arteriosclerosis, myeloma, leukaemia, acquired immunodeficiency

CC syndrome (AIDS), allergies, ovulatory defects, angina, hypertension,

CC stroke, Alzheimer's disease, epilepsy and Tourette's disorder. The

CC polynucleotides may be used to detect and quantify gene expression in

CC biopsied tissues where protein expression may be correlated with disease

CC e.g. to determine absence, presence or excess expression of HTMP or to

CC monitor regulation of HTMP expression during therapeutic intervention.

XX

SQ Sequence 2022 BP; 553 A; 494 C; 459 G; 516 T; 0 other;

Query Match 73.6%; Score 16.2; DB 21; Length 2022;

Best Local Similarity 85.7%; Pred. No. 3.le+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAAGGCACTAGCTTCAGTGTG 22

DB 617 CAAGGCACTATCATCAGTGTG 597

RESULT 11

ABL68244/c

ID ABL68244 standard; DNA; 2226 BP.

XX

AC ABL68244;

XX

15-MAY-2002 (first entry)

XX

DE Kidney cancer related gene sequence SEQ ID NO:6581.

XX

KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;

KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;

KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;

XX gene; ds.

OS Homo sapiens.

XX

PN WO200194629-A2.

XX

PD 13-DEC-2001.

XX

PF 30-MAY-2001; 2001WO-US10838.

XX

PR 05-JUN-2000; 2000US-209473P.

PR 05-JUN-2000; 2000US-209531P.

PR 18-SEP-2000; 2000US-233133P.

PR 18-SEP-2000; 2000US-233617P.

PR 20-SEP-2000; 2000US-234009P.

PR 20-SEP-2000; 2000US-234034P.

PR 20-SEP-2000; 2000US-234052P.

PR 22-SEP-2000; 2000US-234509P.

PR 22-SEP-2000; 2000US-234567P.

PR 25-SEP-2000; 2000US-234923P.

PR 25-SEP-2000; 2000US-234924P.

PR 25-SEP-2000; 2000US-235077P.

PR 25-SEP-2000; 2000US-235082P.

PR 25-SEP-2000; 2000US-235134P.

PR 25-SEP-2000; 2000US-235280P.

PR 26-SEP-2000; 2000US-235637P.

PR 26-SEP-2000; 2000US-235638P.

PR 27-SEP-2000; 2000US-235711P.

PR 27-SEP-2000; 2000US-235720P.

PR 27-SEP-2000; 2000US-235840P.

PR 27-SEP-2000; 2000US-235863P.

PR 28-SEP-2000; 2000US-236028P.

PR 28-SEP-2000; 2000US-236032P.

PR 28-SEP-2000; 2000US-236033P.

PR 28-SEP-2000; 2000US-236034P.

PR 28-SEP-2000; 2000US-236109P.

PR 28-SEP-2000; 2000US-236111P.

PR 29-SEP-2000; 2000US-236842P.

PR 29-SEP-2000; 2000US-236891P.

PR 02-OCT-2000; 2000US-237172P.

PR 02-OCT-2000; 2000US-237173P.

PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.

PR 02-OCT-2000; 2000US-237295P.

PR 02-OCT-2000; 2000US-237316P.

PR 03-OCT-2000; 2000US-237425P.

PR 03-OCT-2000; 2000US-237598P.

PR 03-OCT-2000; 2000US-237604P.

PR 03-OCT-2000; 2000US-237606P.

PR 03-OCT-2000; 2000US-237608P.

PR 01-NOV-2000; 2000US-244867P.

PR 01-NOV-2000; 2000US-245084P.

XX (AVAL-) AVALON PHARM.

FA Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;

XX Soppet DR, Weaver Z;

PI WPI; 2002-188264/24.

DR

XX

Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity and determining a change in expression of a gene of a signature gene set

Claim 1; SEQ ID 6581; 44pp; English.

The present invention describes a method (M1) for screening for an anti-neoplastic agent. The method involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, determining a change in expression of at least one gene (I) of a signature gene set, where (I) comprises a sequence (S) selected from 8447 sequences (given in ABL61664 to ABL70110), or is at least 95% identical to (S), where a change in expression is indicative of anti-neoplastic activity. (I) has cytostatic activity and can be used in gene therapy. M1 can be used for screening an anti-neoplastic agent, and can be used for producing a product which is the data collected with respect to the anti-neoplastic agent as a result of M1, and the data is sufficient to convey the chemical structure and/or properties of the agent. M1 can be used in the treatment of cancer such as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney, prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms' tumour.

Sequence 2226 BP; 412 A; 686 C; 666 G; 462 T; 0 other;

Query Match 73.6%; Score 16.2; DB 24; Length 2226;

Best Local Similarity 85.7%; Pred. No. 3.le+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCAAAGCACTAGCTTCAGTGT 21

DB 1166 GCAGAGCACCAGCTGCAGTGT 1146

RESULT 12

ABL68864/c

ID ABL68864 standard; DNA; 2226 BP.

XX

AC ABL68864;

XX





Db 242 CAAGCACTATCATCAGTGTG 222

Search completed: July 8, 2003, 02:18:58  
Job time : 140.224 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 30.3707 seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-11

Perfect score: 22

Sequence: 1 gcaagcactagcttcagtggtg 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Issued\_Patents\_NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/6C\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.4	74.5	563	4	US-09-242-901-14
2	15.8	71.8	1122	1	US-08-625-876-1
3	15.8	71.8	1122	1	US-08-625-876-3
4	15.8	71.8	1122	1	US-08-625-876-5
5	15.8	71.8	1825	4	US-09-071-035-463
6	15.8	71.8	6168	4	US-09-071-035-457
7	15.8	71.8	6168	4	US-09-071-035-461
8	15.8	71.8	6168	4	US-09-071-035-465
9	15.2	69.1	941	2	US-08-627-610-11
10	15.2	69.1	941	2	US-08-384-106A-6
11	15.2	69.1	941	4	US-09-240-906-3
12	15.2	69.1	941	5	PCT-US96-01643-6
13	15.2	69.1	1001	4	US-09-641-638-601
14	15.2	69.1	1010	5	PCT-US96-05253-3
15	15.2	69.1	2202	4	US-09-388-743-1
16	14.8	67.3	296	2	US-08-716-942-20
17	14.8	67.3	296	4	US-09-130-337A-20
18	14.8	67.3	613	4	US-08-998-416-205
19	14.8	67.3	2676	1	US-08-525-596B-11
20	14.8	67.3	2676	3	US-09-177-860A-11
21	14.8	67.3	2676	3	US-08-891-789B-5
22	14.8	67.3	2676	4	US-09-378-238-11
23	14.8	67.3	2676	4	US-09-451-501-11
24	14.8	67.3	2848	3	US-08-539-205A-3
25	14.8	67.3	168575	4	US-09-426-290-1
26	14.8	67.3	168575	4	US-09-426-290-1
27	14.6	66.4	243	2	US-08-505-218-7

c 28	14.6	66.4	568	2	US-08-505-218-1	Sequence 1, Appl
c 29	14.6	66.4	568	2	US-08-505-218-12	Sequence 12, Appl
c 30	14.6	66.4	637	4	US-08-861-774E-33	Sequence 33, Appl
c 31	14.6	66.4	1114	4	US-09-152-060-41	Sequence 41, Appl
c 32	14.6	66.4	1123	4	US-09-152-060-15	Sequence 15, Appl
c 33	14.6	66.4	1192	4	US-09-318-448-18	Sequence 18, Appl
c 34	14.6	66.4	2016	1	US-08-294-872-1	Sequence 1, Appl
c 35	14.6	66.4	2016	5	PCT-US95-09823-1	Sequence 1, Appl
c 36	14.6	66.4	2352	4	US-08-997-251-3	Sequence 3, Appl
c 37	14.6	66.4	3290	2	US-08-729-955A-1	Sequence 1, Appl
c 38	14.6	66.4	3539	4	US-08-853-948B-1	Sequence 1, Appl
c 39	14.6	66.4	5362	2	US-08-853-310-3	Sequence 3, Appl
c 40	14.4	65.5	1980	4	US-09-134-001C-2700	Sequence 2700, Ap
c 41	14.4	65.5	2525	4	US-09-342-648-1	Sequence 1, Appl
c 42	14.4	65.5	8298	5	PCT-US93-03076-1	Sequence 1, Appl
c 43	14.2	64.5	166	1	US-07-807-043B-16	Sequence 16, Appl
c 44	14.2	64.5	210	4	US-08-847-065-14	Sequence 14, Appl
c 45	14.2	64.5	512	4	US-09-527-223-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-242-901-14  
; Sequence 14, Application US/09242901  
; Patent No. 6322783  
; GENERAL INFORMATION:  
; APPLICANT: TAKAHASHI, SEISHI  
; TITLE OF INVENTION: NOVEL BACTERIOPHAGES, METHOD FOR SCREENING SAME AND  
; TITLE OF INVENTION: BACTERICIDAL COMPOSITIONS USING SAME, AND DETECTION  
; TITLE OF INVENTION: KITS USING SAME  
; FILE REFERENCE: 3000-0001  
; CURRENT APPLICATION NUMBER: US/09/242,901  
; EARLIER FILING DATE: 1999-02-26  
; EARLIER APPLICATION NUMBER: PCT/JP97/02957  
; EARLIER FILING DATE: 1997-08-26  
; EARLIER APPLICATION NUMBER: 8/261132  
; EARLIER FILING DATE: 1996-08-26  
; EARLIER APPLICATION NUMBER: 9/135716  
; EARLIER FILING DATE: 1997-04-19  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 563  
; TYPE: DNA  
; ORGANISM: Bacteriophage  
; FEATURE:  
; OTHER INFORMATION: "n" at various positions throughout the sequence may be  
; OTHER INFORMATION: a, t, g, c other or unknown  
US-09-242-901-14

Query Match 74.5%; Score 16.4; DB 4; Length 563;  
Best Local Similarity 89.5%; Pred. No. 22;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AAGCACTAGCTTCAGTGTG 22  
|||||  
Db 328 AAGCACTAGCTTCAGTGTG 346

RESULT 2  
US-08-625-876-1  
; Sequence 1, Application US/08625876  
; Patent No. 5656471  
; GENERAL INFORMATION:  
; APPLICANT: MINAGAWA, Hirotaka  
; APPLICANT: NAKAYAMA, No. 5656471iuyuki  
; APPLICANT: NAKAMOTO, Shinya  
; TITLE OF INVENTION: LACTATE OXIDASE WITH AN IMPROVED THERMAL  
; TITLE OF INVENTION: STABILITY AND GENE OF THE SAME  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release.#1.0.  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/625,876-1  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-95947  
FILING DATE: 30-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-146186  
FILING DATE: 13-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: PF-1612  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-2297  
TELEFAX: (703) 695-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1122 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1122  
US-08-625-876-1

Query Match 71.8%; Score 15.8; DB 1; Length 1122;  
Best Local Similarity 89.5%; Pred. No. 49;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CAAAGCACTAGCTTCAGTG 20  
||||| ||||| ||||| |  
Db 918 CAAAGCGCTAGCTTCAGGG 936

RESULT 3  
US-08-625-876-3  
Sequence 3, Application US/08625876  
Patent No. 5656471  
GENERAL INFORMATION:  
APPLICANT: MINAGAWA, Hirotaka  
APPLICANT: NAKAYAMA, No. 5656471lyuk1  
APPLICANT: NAKAMOTO, Shinya  
TITLE OF INVENTION: LACTATE OXIDASE WITH AN IMPROVED THERMAL  
TITLE OF INVENTION: STABILITY AND GENE OF THE SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

```

CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/625,876
  FILING DATE: 01-APR-1996
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 7-95947
    FILING DATE: 30-MAR-1995
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: JP 7-146186
    FILING DATE: 13-JUN-1995
  ATTORNEY/AGENT INFORMATION:
    NAME: PATCH, Andrew J
    REGISTRATION NUMBER: 32,925
    REFERENCE/DOCKET NUMBER: PR-1612
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (703) 521-2297
    TELEFAX: (703) 685-0573
    TELEX: 248425 EMBON
  INFORMATION FOR SEQ ID NO: 3:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 1122 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: other nucleic acid
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 1...1122
US-08-625-876-3

```

Query Match	71.8%;	Score 15.8;	DB 1;	Length 112;
Best Local Similarity	89.5%;	Pred. No. 49;		
Matches 17;	Conservative 0;	Mismatches 2;	Indels 0	
Qy	2	CAAAGCCTAGCTTCAGTG	20	
Db	918	CAAAAGCGCTAGCTTCCAGG	936	

RESULT 4  
US-08-625-876-5  
; Sequence 5, Application US/08625876  
; Patent No. 5656471  
; GENERAL INFORMATION:  
; APPLICANT: MINAGAWA, Hirotaka  
; APPLICANT: NAKAYAMA, No. 56564711yuki  
; APPLICANT: NAKAMOTO, Shinya  
; TITLE OF INVENTION: LACTATE OXIDASE WITH AN IMPROVED THERMAL  
; TITLE OF INVENTION: STABILITY AND GENE OF THE SAME  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: YOUNG & THOMPSON  
; STREET: 745 South 23rd Street  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/625,876  
; FILING DATE: 01-APR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-95947  
; FILING DATE: 30-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP-7-146186  
; FILING DATE: 13-JUN-1995  
; ATTORNEY/AGENT INFORMATION:

NAME: PATCH, Andrew J  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: PF-1612  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-2297  
TELEFAX: (703) 685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1122 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1122  
US-08-625-876-5

Query Match 71.8%; Score 15.8; DB 1; Length 1122;  
Best Local Similarity 89.5%; Pred. No. 49;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CAAGCAGCTAGCTTCAGTG 20  
Db 918 CAAGCGCTAGCTTCAGGG 936

RESULT 5  
US-09-071-035-463/c  
; Sequence 463, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 463:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1825 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-071-035-463

Query Match 71.8%; Score 15.8; DB 4; Length 1825;  
Best Local Similarity 89.5%; Pred. No. 53;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAGCACTAGCTTCAGTGT 21  
Db 1635 AAAGCACTTGTTTCAGTGT 1617

RESULT 6  
US-09-071-035-457/c  
; Sequence 457, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 457:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6168 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-09-071-035-457

Query Match 71.8%; Score 15.8; DB 4; Length 6168;  
Best Local Similarity 89.5%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AAAGCACTAGCTTCAGTGT 21  
Db 4151 AAAGCACTTGTTTCAGTGT 4133

RESULT 7  
US-09-071-035-461/c  
; Sequence 461, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
US-09-071-035-461

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 461:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6168 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-071-035-461

Query Match 71.8%; Score 15.8; DB 4; Length 6168;  
Best Local Similarity 89.5%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAGCACTAGCTTCAGTGT 21  
||||| | | | | | | | | |  
Db 4151 AAAGCACTGGTTCAGTGT 4133

RESULT 8  
US-09-071-035-465/c  
Sequence 465, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 465:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6168 base pairs  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
US-09-071-035-465  
Query Match 71.8%; Score 15.8; DB 4; Length 6168;  
Best Local Similarity 89.5%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAGCACTAGCTTCAGTGT 21  
||||| | | | | | | | | |  
Db 4151 AAAGCACTGGTTCAGTGT 4133

RESULT 9  
US-08-627-610-11  
Sequence 11, Application US/08627610  
Patent No. 5919997  
GENERAL INFORMATION:  
APPLICANT: Beach, David H.  
APPLICANT: Serrano, Manuel  
APPLICANT: Depinho, Ronald A.  
TITLE OF INVENTION: Transgenic Animals Having Modified Cell-Cycle  
TITLE OF INVENTION: Regulation  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/627,610  
FILING DATE: 04-APR-1996  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: CSI-001CP6  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 134..631  
US-08-627-610-11  
Query Match 69.1%; Score 15.2; DB 2; Length 941;  
Best Local Similarity 85.0%; Pred. No. 96;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGCACTAGCTTCAGTGT 22  
||||| | | | | | | | | |  
Db 904 AAAGCCCTAGCCTCAGAGTG 923

RESULT 10  
US-08-384-106A-6  
Sequence 6, Application US/08384106A  
Patent No. 6033847  
GENERAL INFORMATION:

APPLICANT: Sherr Ph.D., Charles J.  
APPLICANT: Downing M.D., James  
APPLICANT: Hirai Ph.D., Hiroshi  
APPLICANT: Okuda, Tsukasa  
TITLE OF INVENTION: InK4c-p18 and InK4d-p19, Inhibitors of  
TITLE OF INVENTION: Cyclin Dependent Kinases CDK4 and CDK6, and Uses Thereof  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX  
STREET: 1100 New York Ave., N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/384,106A  
FILING DATE: 06-FEB-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, Samuel L.  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 0656.0500000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-384-106A-6

Query Match 69.1%; Score 15.2; DB 3; Length 941;  
Best Local Similarity 85.0%; Pred. No. 96;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGTG 22  
||||| ||||| ||||| ||||| |||||  
DB 904 AAAGCCTAGCCTCAGAGTG 923

RESULT 11  
US-09-240-906-3  
Sequence 3, Application US/09240906  
Patent No. 6245965  
GENERAL INFORMATION:  
APPLICANT: ROUSSEL, MARTINE F.  
APPLICANT: ZINDY, RICHARD  
APPLICANT: ZINDY, FREDERIQUE  
APPLICANT: CUNNINGHAM, JUSTINE  
TITLE OF INVENTION: CELLS THAT LACK P19INK4D AND P27KIP1 ACTIVITY AND  
TITLE OF INVENTION: METHODS OF USE THEREOF  
FILE REFERENCE: 1340-1-025  
CURRENT APPLICATION NUMBER: US/09/240,906  
CURRENT FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 3  
LENGTH: 941  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-240-906-3

Query Match 69.1%; Score 15.2; DB 4; Length 941;  
Best Local Similarity 85.0%; Pred. No. 96;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGTG 22  
||||| ||||| ||||| ||||| |||||  
DB 904 AAAGCCTAGCCTCAGAGTG 923

RESULT 12  
PCT-US96-01643-6  
Sequence 6, Application PC/TUS9601643.  
GENERAL INFORMATION:  
APPLICANT: St. Jude Children's Research Hospital  
TITLE OF INVENTION: InK4c-p18 and InK4d-p19, Inhibitors of  
TITLE OF INVENTION: Cyclin Dependent Kinases CDK4 and CDK6, and Uses Thereof  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.  
STREET: 1100 New York Ave., N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/01643  
FILING DATE: 06-FEB-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/384,106  
FILING DATE: 06-FEB-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, Samuel L.  
REGISTRATION NUMBER: 30,353  
REFERENCE/DOCKET NUMBER: 0656.050PC01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 941 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
STRAIN: C57 Black Kaplan  
CELL LINE: V13  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: p19  
PCT-US96-01643-6

Query Match 69.1%; Score 15.2; DB 5; Length 941;  
Best Local Similarity 85.0%; Pred. No. 96;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGTG 22  
||||| ||||| ||||| ||||| |||||  
DB 904 AAAGCCTAGCCTCAGAGTG 923

RESULT 13  
US-09-641-638-601  
Sequence 601, Application US/09641638  
Patent No. 6432648  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Chumakov, Ilya

```
;; APPLICANT: Cohen, Annick
;; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;; FILE REFERENCE: GENSET.051CPI
;; CURRENT APPLICATION NUMBER: US/09/641,638
;; PRIOR FILING DATE: 2000-08-16
;; PRIOR APPLICATION NUMBER: US 09/502,330
;; PRIOR FILING DATE: 2000-02-11
;; PRIOR APPLICATION NUMBER: US 60/133,200
;; PRIOR FILING DATE: 1999-05-07
;; PRIOR APPLICATION NUMBER: US-09/275,267
;; PRIOR FILING DATE: 1999-03-23
;; PRIOR APPLICATION NUMBER: US 60/119,917
;; PRIOR FILING DATE: 1999-02-12
;; NUMBER OF SEQ ID NOS: 1304
;; SOFTWARE: Patent.pm
;; SEQ ID NO 601
;; LENGTH: 1001
;; TYPE: DNA
;; ORGANISM: Homo Sapiens
;; FEATURE:
;; NAME/KEY: allele
;; LOCATION: 501
;; OTHER INFORMATION: 10-387-371 : polymorphic base C or T
;; NAME/KEY: misc_binding
;; LOCATION: 481..500
;; OTHER INFORMATION: 10-387-371.misl, potential
;; NAME/KEY: misc_binding
;; LOCATION: 502..521
;; OTHER INFORMATION: 10-387-371.mis2, potential complement
;; NAME/KEY: primer_bind
;; LOCATION: 131..149
;; OTHER INFORMATION: upstream amplification primer
;; NAME/KEY: primer_bind
;; LOCATION: 535..552
;; OTHER INFORMATION: downstream amplification primer, complement
;; NAME/KEY: misc_binding
;; LOCATION: 489..513
;; OTHER INFORMATION: 10-387-371 potential probe
US-09-641-638-601

Query Match 69.1%; Score 15.2; DB 4; Length 1001;
Best Local Similarity 85.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGTG 22
| | | | | | | | | | | | | | | |
Db 770 AGAGCAGCTAGCTTCATAGTG 789

RESULT 14
PCT-US96-05252-3
Sequence 3, Application PC/TUS9605252
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: p19: A Cell Cycle Inhibitor
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew
CITY: One Market Plaza, Steuart Street Tower
STATE: San Francisco
COUNTRY: USA
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: PCT/US96/05252
FILING DATE: Not yet assigned
CLASSIFICATION:
```

```
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/425,093
;; FILING DATE: 17-APR-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Weber, Ellen L.
;; REGISTRATION NUMBER: 32,762
;; REFERENCE/DOCKET NUMBER: 02307B-059910PC
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 543-9600
;; TELEFAX: (415) 543-5043
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1010 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 28..525
;; OTHER INFORMATION: /product= "murine p19"
PCT-US96-05252-3

Query Match 69.1%; Score 15.2; DB 5; Length 1010;
Best Local Similarity 85.0%; Pred. No. 97;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGTG 22
| | | | | | | | | | | | | | | |
Db 802 AAAGCCTAGCCTCAGAGTG 821

RESULT 15
US-09-388-743-1/c
Sequence 1, Application US/09388743
Patent No. 6423886
GENERAL INFORMATION:
APPLICANT: Singletary, George
APPLICANT: Zhou, Lan
TITLE OF INVENTION: No. 6423886el Starch Synthase Polynucleotides and Their
FILE REFERENCE: 1144
CURRENT APPLICATION NUMBER: US/09/388,743
CURRENT FILING DATE: 1999-09-02
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 2202
TYPE: DNA
ORGANISM: Curcuma zedoaria
FEATURE:
NAME/KEY: CDS
LOCATION: (130)...(1974)
US-09-388-743-1

Query Match 69.1%; Score 15.2; DB 4; Length 2202;
Best Local Similarity 85.0%; Pred. No. 1.le+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAAAGCAGCTAGCTTCAGTGT 21
| | | | | | | | | | | | | | | |
Db 207 CAATCAGCTAGCTCCATGT 188

Search completed: July 8, 2003, 09:31:46
Job time : 32.4207 secs
```

GenCore version 5.1.6.  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 127.063 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-11

Perfect score: 22

Sequence: 1 gcaagcactagcttcagtgtg 22

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	16.8	76.4	331	10	US-09-917-800A-461
2	16.8	76.4	7444	9	US-09-909-567B-37
3	16.8	76.4	7879	9	US-10-198-846-9851
4	16.2	73.6	219	9	US-10-053-886-4
5	16.2	73.6	295	10	US-09-294-093B-2504
6	16.2	73.6	682	9	US-10-043-487-24
7	16.2	73.6	1067	9	US-10-071-766-30
8	16.2	73.6	2274	9	US-10-029-180-91
9	16.2	73.6	25701	9	US-09-764-891-9766
10	16.2	73.6	25758	9	US-09-764-891-9765
11	15.8	71.8	522	9	US-10-198-846-12317
12	15.8	71.8	872	9	US-10-198-846-7129
13	15.8	71.8	1947	10	US-09-864-761-4452
14	15.8	71.8	2298	9	US-10-217-357-3
15	15.8	71.8	2298	10	US-09-975-326-3
16	15.8	71.8	2301	9	US-09-934-406-3
17	15.8	71.8	2301	9	US-10-217-357-1
18	15.8	71.8	2301	10	US-09-975-326-1
19	15.8	71.8	3552	9	US-09-934-406-1

c	20	15.8	71.8	27360	10	US-09-070-927A-164
	21	15.8	71.8	119596	9	US-10-270-336-3
	22	15.6	70.9	421	9	US-09-918-995-34956
c	23	15.6	70.9	499	10	US-09-864-761-44
	24	15.6	70.9	573	10	US-09-764-877-790
	25	15.6	70.9	573	10	US-09-764-877-3579
	26	15.6	70.9	573	10	US-09-764-877-3580
	27	15.6	70.9	573	10	US-09-764-877-3581
	28	15.6	70.9	573	10	US-09-764-877-3582
	29	15.6	70.9	785	9	US-10-198-846-4207
c	30	15.6	70.9	844	9	US-10-198-846-4209
	31	15.6	70.9	855	9	US-10-198-846-4041
	32	15.6	70.9	860	9	US-10-198-846-4087
c	33	15.6	70.9	908	9	US-10-198-846-12073
	34	15.6	70.9	939	9	US-09-764-881-61
	35	15.6	70.9	1057	9	US-10-106-698-741
	36	15.6	70.9	2438	9	US-10-232-484-3
	37	15.6	70.9	5207	9	US-10-098-841-59
c	38	15.6	70.9	5230	9	US-10-213-948-8
	39	15.6	70.9	5261	9	US-10-098-841-58
	40	15.6	70.9	32185	10	US-09-764-877-3171
	41	15.4	70.0	1197	9	US-10-198-846-10718
	42	15.4	70.0	335913	9	US-09-754-853A-2
	43	15.4	70.0	335913	9	US-09-754-853A-3
	44	15.2	69.1	25	9	US-10-098-263B-121417
	45	15.2	69.1	223	10	US-09-764-877-534

#### ALIGNMENTS

#### RESULT 1

US-09-917-800A-461  
; Sequence 461, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917 800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 461  
; LENGTH: 331  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA998276  
; US-09-917-800A-461

Sequence 164, App  
Sequence 3, Appli  
Sequence 34956, A  
Sequence 44, Appl  
Sequence 790, App  
Sequence 3579, Ap  
Sequence 3580, Ap  
Sequence 3581, Ap  
Sequence 3582, Ap  
Sequence 4207, Ap  
Sequence 4209, Ap  
Sequence 4041, Ap  
Sequence 4087, Ap  
Sequence 12073, A  
Sequence 61, Appl  
Sequence 741, App  
Sequence 3, Appli  
Sequence 59, Appl  
Sequence 8, Appli  
Sequence 58, Appl  
Sequence 3171, Ap  
Sequence 10718, A  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 121417,  
Sequence 534, App

```

; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature

```

APPLICANT: SHEPHERD, R. L.

TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL

OTHER INFORMATION: Incyte ID No. US20010051335A1 700346213H1  
NAME/KEY: unsure  
LOCATION: 273  
OTHER INFORMATION: a, t, c, g, or other  
US-09-294-093B-2504

Query Match 73.6%; Score 16.2; DB 10; Length 295;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAAAGCACTAGCTTCAGTGTG 22  
||||| ||| ||||| |||  
Db 93 CAAAGTACTGCTTCAGAGTG 113

RESULT 6  
US-10-043-487-24/c  
Sequence 24, Application US/10043487.  
Publication No. US20030055220A1

GENERAL INFORMATION:  
APPLICANT: HYBRIGENICS  
APPLICANT: PIERRE, LEGRIN  
TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptides  
FILE REFERENCE: B4778A  
CURRENT APPLICATION NUMBER: US/10/043,487  
PRIOR FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/261,130  
PRIOR FILING DATE: 2001-01-12  
NUMBER OF SEQ ID NOS: 561  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 24  
LENGTH: 682  
TYPE: DNA  
ORGANISM: Shigella Flexneri  
US-10-043-487-24

Query Match 73.6%; Score 16.2; DB 9; Length 682;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CAAAGCACTAGCTTCAGTGTG 22  
||||| ||| ||||| |||  
Db 493 CAAGGCATCATCATCAGTGTG 473

RESULT 7  
US-10-071-766-30  
Sequence 30, Application US/10071766  
Publication No. US20020192678A1  
GENERAL INFORMATION:  
APPLICANT: Huel-Mei Chen  
TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE  
FILE REFERENCE: PA-0043 US  
CURRENT APPLICATION NUMBER: US/10/071,766  
CURRENT FILING DATE: 2002-02-07  
NUMBER OF SEQ ID NOS: 144  
SOFTWARE: PERL Program  
SEQ ID NO 30  
LENGTH: 1067  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US20020192678A1 1100140.10  
NAME/KEY: unsure  
LOCATION: 717-759  
OTHER INFORMATION: a, t, c, g, or other  
US-10-071-766-30

Query Match 73.6%; Score 16.2; DB 9; Length 1067;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCAAAGCACTAGCTTCAGTGT 21  
||||| ||||| ||||| |||  
Db 951 GCAAATACTAGCTTCAGTAT 971

RESULT 8  
US-10-029-180-91/c  
Sequence 91, Application US/10029180  
Publication No. US20020182708A1  
GENERAL INFORMATION:  
APPLICANT: Cali, Brian M.  
APPLICANT: Holtzman, Doug  
APPLICANT: Madden, Kevin T.  
APPLICANT: Milna, G. Todd  
APPLICANT: Sherman, Amir  
APPLICANT: Silva, Jeffrey C.  
APPLICANT: Trueheart, Josh  
APPLICANT: Zhang, Lixin  
TITLE OF INVENTION: No. US20020182708A1el Regulators of Fungal Gene Expression  
FILE REFERENCE: MIC-004  
CURRENT APPLICATION NUMBER: US/10/029,180  
CURRENT FILING DATE: 2001-12-22  
PRIOR APPLICATION NUMBER: US 60/257,431  
PRIOR FILING DATE: 2000-12-22  
NUMBER OF SEQ ID NOS: 138  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 91  
LENGTH: 2274  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fungal gene  
US-10-029-180-91

Query Match 73.6%; Score 16.2; DB 9; Length 2274;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCAAAGCACTAGCTTCAGTGT 21  
||||| ||||| ||||| |||  
Db 210 GCAGAGCCTAGCTCCAGTGT 190

RESULT 9  
US-09-764-891-9766/c  
Sequence 9766, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9766  
LENGTH: 25701  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-9766

Query Match 73.6%; Score 16.2; DB 9; Length 25701;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCAAAGCACTAGCTTCAGTGT 21  
||||| ||||| ||||| |||  
Db 24714 GCAAAGCACTACTTCACTGTCT 24694

RESULT 10

US-09-764-891-9765/c  
; Sequence 9765, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 9765  
; LENGTH: 25758  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-9765

Query Match 73.6%; Score 16.2; DB 9; Length 25758;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCAAGCAGCTAGCTTCAGTGT 21  
|||||  
DB 24771 GCAAGCAGCTAGCTTCAGTGT 24751

RESULT 11  
US-10-198-846-12317  
; Sequence 12317, Application US/10198846  
; Publication No. US20030099974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steimann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12317  
; LENGTH: 522  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-846-12317

Query Match 71.8%; Score 15.8; DB 9; Length 522;  
Best Local Similarity 89.5%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGT 21  
|||||  
DB 291 AAATCACTAGCTTCTGTGT 309

RESULT 12  
US-10-198-846-7129  
; Sequence 7129, Application US/10198846  
; Publication No. US20030099974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steimann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF BREAST CANCER

FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10/198,846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7129  
; LENGTH: 872  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 555, 558, 583, 618, 675, 681, 706, 712, 733, 734, 736, 774,  
; LOCATION: 785, 789, 796, 803, 814, 842, 859.  
; OTHER INFORMATION: n = A,T,C or G  
US-10-198-846-7129

Query Match 71.8%; Score 15.8; DB 9; Length 872;  
Best Local Similarity 89.5%; Pred. No. 2.1e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AAAGCAGCTAGCTTCAGTGT 21  
|||||  
DB 397 AAATCACTAGCTTCTGTGT 415

RESULT 13  
US-09-864-761-4452  
; Sequence 4452, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemlica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21

;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 4452  
;; LENGTH: 1947  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC003682.1  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5  
US-09-864-761-4452

Query Match 71.8%; Score 15.8; DB 10; Length 1947;  
Best Local Similarity 89.5%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAAAGCAGCTAGCTTCAGT 19  
||||| | |||||  
DB 1417 GCAAAGCTCCAGCTTCAGT 1435

RESULT 14  
US-10-217-357-3  
;; Sequence 3, Application US/10217357  
;; Publication No. US20030023063A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Turner, C. Alexander Jr.  
;; APPLICANT: Mathur, Brian  
;; TITLE OF INVENTION: No. US20030023063A1el Human Kinases and Polynucleotides Encoding  
;; FILE OF INVENTION: Same  
;; FILE REFERENCE: LEX-0254-USA  
;; CURRENT APPLICATION NUMBER: US/10/217,357  
;; CURRENT FILING DATE: 2002-08-09  
;; PRIOR APPLICATION NUMBER: US/09/975,326  
;; PRIOR FILING DATE: 2001-10-11  
;; PRIOR APPLICATION NUMBER: US 60/239,821  
;; PRIOR FILING DATE: 2000-10-12  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3  
;; LENGTH: 2298  
;; TYPE: DNA  
;; ORGANISM: homo sapiens  
US-10-217-357-3

Query Match 71.8%; Score 15.8; DB 9; Length 2298;  
Best Local Similarity 89.5%; Pred. No. 2.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAAAGCAGCTAGCTTCAGT 19  
||||| | |||||  
DB 944 GCAAATCACCAGCTTCAGT 962

RESULT 15  
US-09-975-326-3  
;; Sequence 3, Application US/09975326  
;; Patent No. US20020128458A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Turner, C. Alexander Jr.  
;; APPLICANT: Mathur, Brian  
;; TITLE OF INVENTION: No. US20020128458A1el Human Kinases and Polynucleotides Encoding

;; FILE REFERENCE: LEX-0254-USA  
;; CURRENT APPLICATION NUMBER: US/09/975,326  
;; CURRENT FILING DATE: 2001-10-11  
;; PRIOR APPLICATION NUMBER: US 60/239,821  
;; PRIOR FILING DATE: 2000-10-12  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3  
;; LENGTH: 2298  
;; TYPE: DNA  
;; ORGANISM: homo sapiens  
US-09-975-326-3

Query Match 71.8%; Score 15.8; DB 10; Length 2298;  
Best Local Similarity 89.5%; Pred. No. 2.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAAAGCAGCTAGCTTCAGT 19  
||||| | |||||  
DB 944 GCAAATCACCAGCTTCAGT 962

Search completed: July 9, 2003, 02:22:05  
Job time : 128.113 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1115.35 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-11

Perfect score: 22

Sequence: 1 gcaagcactagcttcagtgtg 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba.\*

2: em\_esthum.\*

3: em\_estin.\*

4: em\_estmu.\*

5: em\_estov.\*

6: em\_estpl.\*

7: em\_estro.\*

8: em\_htc.\*

9: gb\_est1.\*

10: gb\_est2.\*

11: gb\_htc.\*

12: gb\_est3.\*

13: gb\_est4.\*

14: gb\_est5.\*

15: em\_estfun.\*

16: em\_estom.\*

17: gb\_gss.\*

18: em\_gss\_hum.\*

19: em\_gss\_inv.\*

20: em\_gss\_pln.\*

21: em\_gss\_vrt.\*

22: em\_gss\_fun.\*

23: em\_gss\_mam.\*

24: em\_gss\_mus.\*

25: em\_gss\_other.\*

26: em\_gss\_pro.\*

27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	18	81.8	684	10	AV383817
c 2	17.8	80.9	459	17	BH770100
c 3	17.8	80.9	526	17	A2443237
c 4	17.8	80.9	590	17	BH477117
c 5	17.4	79.1	473	17	A2915638
c 6	17.4	79.1	768	17	BH117517

7	17.4	79.1	822	12	BG676524
8	17.2	78.2	165	10	BE061457
9	17.2	78.2	229	17	BH058084
10	17.2	78.2	256	12	BG095858
11	17.2	78.2	277	9	AV270745
12	17.2	78.2	402	17	AQ586256
13	17.2	78.2	427	17	BH121214
14	17.2	78.2	454	13	BH536125
15	17.2	78.2	524	10	BE341990
16	17.2	78.2	569	12	BG081637
17	17.2	78.2	598	17	BH533618
18	17.2	78.2	610	17	AQ656158
19	17.2	78.2	613	17	BH351842
20	17.2	78.2	690	12	BG589769
21	17.2	78.2	762	17	AZ703949
22	17.2	78.2	968	12	BH138199
23	17.2	78.2	1101	17	CNS017M8
24	17.2	77.3	832	9	AL579993
25	16.8	76.4	274	9	AV212423
26	16.8	76.4	287	9	AI473268
27	16.8	76.4	289	9	AV099156
28	16.8	76.4	331	9	AA998276
29	16.8	76.4	378	12	BG466040
30	16.8	76.4	398	10	AW285513
31	16.8	76.4	400	10	AW285496
32	16.8	76.4	406	10	AW680420
33	16.8	76.4	422	9	AA754067
34	16.8	76.4	428	10	AW924299
35	16.8	76.4	435	10	AV669321
36	16.8	76.4	449	10	AW922919
37	16.8	76.4	454	12	BG101962
38	16.8	76.4	464	12	BG241541
39	16.8	76.4	467	9	AA479142
40	16.8	76.4	493	12	BG048839
41	16.8	76.4	499	17	AQ333196
42	16.8	76.4	506	13	BH328697
43	16.8	76.4	520	17	AQ565228
44	16.8	76.4	542	17	AQ685473
45	16.8	76.4	563	10	BE363430

# ALIGNMENTS

RESULT 1  
AV383817/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

AV383817 684 bp mRNA linear EST 27-OCT-1999  
AV383817 Halocynthia roretzi Fertilized egg Halocynthia roretzi  
CDNA clone 007E18\_5', mRNA sequence.  
AV383817  
GI:6128874  
1 (bases 1 to 684)  
Halocynthia roretzi.  
Halocynthia roretzi  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
Stolidobranchia; Pyruidae; Halocynthia.  
Makabe, K.W.  
Halocynthia roretzi EST  
Unpublished (1999)  
Contact: Kazuhiro W. Makabe  
Department of Zoology, Graduate School of Science  
Kyoto University  
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4095  
Fax: 81-75-705-1113  
Email: kwakabe@sci.kyoto-u.ac.jp.  
Location/Qualifiers  
1.684  
/organism="Halocynthia roretzi"  
/db\_xref="taxon:7729"  
/clone="007E18\_5"  
/clone\_lib="Halocynthia roretzi Fertilized egg"

```

/dev.stage="fertilized egg"
/note="Organ: embryo"
BASE COUNT      219 a   132 c   148 g   182 t      3 others
ORIGIN

Query Match      81.8%; Score 18; DB 10; Length 684;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AAGCACTAGCTTCAGTGT 21
    |||||
Db 442 AAGCACTAGCTTCAGTGT 425

RESULT 2
BH770100/c
LOCUS      BH770100      459 bp      DNA      linear      GSS 20-MAR-2002
DEFINITION BMBAC369H0277_P5U Brugia malayi Genomic Bac Library 3 Brugia malayi
            genomic, DNA sequence.
ACCESSION  BH770100
VERSION     BH770100.1 GI:19567864
KEYWORDS   GSS.
SOURCE     Brugia malayi.
ORGANISM   Brugia malayi
            Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
            Onchocercidae; Brugia..
REFERENCE  1 (bases 1 to 459)
AUTHORS   Whitton,C., Daub,J., Ware,J., Quail,M., Hall,N., Barrell,B., Foster
            J., Guiliano,D., Slatko,B. and Blaxter M.
TITLE     Genome survey sequences from the human parasitic nematode Brugia
            malayi
JOURNAL   Unpublished (2000)
COMMENT   Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            Sequenced from the Brugia malayi BAC library constructed by Claire
            Whitton and Dr Mike Quail. The sequence was generated by The
            Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in
            collaboration with Mark Blaxter, ICAPB, University of Edinburgh,
            Edinburgh, UK.
            Seq primer: T7 (TAATACGACTCATATAGGG)
            Class: BAC ends.
FEATURES             source
    source
    1..459
        /organism="Brugia malayi"
        /strain="FRS"
        /db_xref="taxon:6279"
        /clone_lib="Brugia malayi Genomic Bac Library 3"
        /sex="Mixed (male and female)"
        /tissue_type="whole parasite"
        /dev_stage="microfilaria (L1)"
        /note="vector: pBAC3.6; Site_1: BamH I; Brugia malayi
            genomic DNA was partially cleaved with Sau3A I and size
            fractionated. 7,392 clones were generated with mean insert
            size ~48 kbp. The library was constructed by Claire
            Whitton, Blaxter Nematode Genetics Lab, University of
            Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing
            Unit, The Sanger Centre, Cambridge, UK."
BASE COUNT      133 a   94 c   107 g   125 t
ORIGIN

Query Match      80.9%; Score 17.8; DB 17; Length 459;
Best Local Similarity 90.5%; Pred. No. 7.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCAAGCACTAGCTTCAGTGT 21
    |||||
Db 210 GCAGAGCACTTGCCTTCAGTGT 190

/dev.stage="fertilized egg"
/note="Organ: embryo"
BASE COUNT      219 a   132 c   148 g   182 t      3 others
ORIGIN

Query Match      81.8%; Score 18; DB 10; Length 684;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 AAGCACTAGCTTCAGTGT 21
    |||||
Db 442 AAGCACTAGCTTCAGTGT 425

RESULT 3
AZ443237/c
LOCUS      1M0237M14R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0237M14 R, DNA sequence.
ACCESSION  AZ443237
VERSION     AZ443237.1 GI:10591023
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
            1 (bases 1 to 526)
REFERENCE  Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0237 row: M column: 14
            Seq primer: CACACGAGAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 526.
FEATURES             Location/Qualifiers
    1..526
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone_lib="UUGC1M0237M14"
        /clone_lib="Mouse 10kb plasmid UUGC1M library"
        /sex="Male"
        /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
        /note="vector: PWD42nv; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource.
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydrodynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel
            electrophoresis. Vector DNA was prepared from a derivative
            of PWD42 (gi|4732114|gb|AF129072.1), a copy-number
            inducible derivative of plasmid R1. The vector was ligated
            with adaptors complementary to the insert adaptors and
            purified. The sheared, adaptor mouse DNA was annealed to
            adaptor vector DNA, and transformed into
            chemically-competent E. coli XL10-Gold (Stratagene) cells
            and selected for ampicillin resistance."
BASE COUNT      174 a   89 c   76 g   187 t
ORIGIN

Query Match      80.9%; Score 17.8; DB 17; Length 526;
Best Local Similarity 90.5%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CAAAGCACTAGCTTCAGTGT 22
    |||||
Db 35 CCAGCACTAGCTTCAGTGT 15

```

## RESULT 4

BH477117  
 LOCUS BH477117 590 bp DNA linear GSS 13-DEC-2001  
 DEFINITION BOHF Brassica oleracea genomic clone BOHFH90, DNA  
 sequence.  
 ACCESSION BH477117  
 VERSION BH477117  
 KEYWORDS BH477117.1 GI:17685228  
 SOURCE GSS.  
 ORGANISM Brassica oleracea.

Brassica oleracea  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.  
 1 (bases 1 to 590)  
 Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.  
 Whole genome shotgun sequencing of Brassica oleracea  
 Unpublished (2001)  
 Other\_GSSs: BOHFH90TF  
 Contact: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 DNA is from a doubled haploid provided by Tom Osborn.  
 Seq primer: TR  
 Class: sheared ends.

## FEATURES

Location/Qualifiers  
 1..590  
 /organism="Brassica oleracea"  
 /strain="T01000DH3"  
 /db\_xref="taxon:3712"  
 /clone="BOHFH90"  
 /cell\_lib="BOHF"  
 /note="Vector: PHOS1; Site\_1: BstXI; 2-3 kb sheared  
 genomic DNA inserted into PHOS1 using BstXI linkers."

BASE COUNT 153 a 124 c 106 g 207 t  
 ORIGIN  
 Query Match 80.9%; Score 17.8; DB 17; Length 590;  
 Best Local Similarity 90.5%; Pred No. 8.8e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 CAAGACACTAGCTTCAGTGTG 22

Db 527 CAATCACTTGTCTCAGTGTG 547

## RESULT 5

AZ915638  
 LOCUS AZ915638 473 bp DNA linear GSS 05-MAR-2001  
 DEFINITION RPCI-24-177N20.TJ RPCI-24 Mus musculus genomic clone RPCI-24-177N20  
 DNA sequence.

ACCESSION AZ915638  
 VERSION AZ915638.1 GI:13234683  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 473)  
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 Unpublished (1999)

Other\_GSSs: RPCI-24-177N20.TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA

## TITLE

## JOURNAL

## COMMENT

Tel: 301 838 0200  
 Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html  
 Plate: 177 row: N column: 20

Seq primer: SP6  
 Class: BAC ends.

## FEATURES

Location/Qualifiers  
 1..473

source

/organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-177N20"  
 /cell\_lib="RPCI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site\_1: BamHI; Site\_2: BamHI;  
 RPCI-24 Mouse BAC library produced by Pieter de Jong. The  
 library was cloned in the pTARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."

BASE COUNT 129 a 81 c 76 g 187 t  
 ORIGIN

Query Match 79.1%; Score 17.4; DB 17; Length 473;  
 Best Local Similarity 94.7%; Pred No. 1.2e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GCAAGCACTAGCTTCAGT 19

Db 388 GCAAGCACTAGCTTCAGT 406

## RESULT 6

BH17517/c

LOCUS BH17517 768 bp DNA linear GSS 19-JUL-2001

DEFINITION RPCI-24-283H21.TJ RPCI-24 Mus musculus genomic clone RPCI-24-283H21

DNA sequence.

ACCESSION BH17517

VERSION BH17517.1 GI:14958860

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 768)

## REFERENCE

## AUTHORS

Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
 Tsegaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P. and Fraser, C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other\_GSSs: RPCI-24-283H21.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (http://www.choi.org/bacpac/orderingframe.htm). BAC end

page: http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html

Plate: 283 row: H column: 21

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1..768

## FEATURES

## source

```

/organism="Mus musculus"
/db_strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-283421"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/notes="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      191 a 134 c 186 g 257 t
ORIGIN
Query Match      79.18; Score 17.4; DB 17; Length 768;
Best Local Similarity 94.78; Pred. No. 1.6e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCAAAGCAGCTAGCTTCAGT 19
|||||
Db 66 GCAAAGCAGCTAGCTTCAGT 48

RESULT 7
BG676524      822 bp mRNA linear EST 01-MAY-2001
LOCUS         602622988f1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4748036 5',
DEFINITION    mRNA sequence.
ACCESSION     BG676524
VERSION       BG676524.1 GI:13907921
KEYWORDS      EST.
SOURCE        Human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 822)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs@mail.nih.gov
              Tissue Procurement: James Cleaver, M.D.
              cDNA Library Preparation: Life Technologies, Inc.
              DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              CDNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: L1AM10598 row: p column: 21
              High quality sequence stop: 740.
              Location/Qualifiers
                1. .822
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4748036"
                  /clone_lib="NCI_CGAP_Skn4"
                  /tissue_type="squamous cell carcinoma"
                  /lab_host="DH10B (TI phage-resistant)"
                  /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                  Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                  Average insert size 1.5kb. Library constructed by Life
                  Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      176 a 241 c 242 g 163 t
ORIGIN
Query Match      79.18; Score 17.4; DB 12; Length 822;
Best Local Similarity 94.78; Pred. No. 1.7e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AAGCAGCTAGCTTCAGTGTG 22
|||||
Db 117 AAGCAGCTAGCTTCAGTGTG 135

```

```

RESULT 8
BE061457/c
LOCUS         RC6-BT0247-240999-002-c05 BT0247 Homo sapiens cDNA, mRNA sequence.
DEFINITION    RC6-BT0247-240999-002-c05 BT0247 Homo sapiens cDNA, mRNA sequence.
ACCESSION     BE061457
VERSION       BE061457.1 GI:8406107
KEYWORDS      EST.
SOURCE        Human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 165)
AUTHORS      Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
              Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
              Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bal, G.S., Simpson, D.H.,
              Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
              M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
              Simpson, A.J.
              Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
              Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
              20202663
              Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?i=st2-RC6-BT0247-240
              999-002-c05&t3=1999-09-24&t4=1)
              Seq primer: puc 18 forward
              High quality sequence start: 21
              High quality sequence stop: 165.
              Location/Qualifiers
                1. .165
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_lib="BT0247"
                  /dev_stage="Adult"
                  /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
                  SmaI; A mini-library was made by cloning products derived
                  from ORESTES PCR (U.S. Letters Patent application No. 196
                  ,716." Ludwig Institute for Cancer Research) profiles
                  into the pUC 18 vector. Reverse transcription of tissue
                  mRNA and cDNA amplification were performed under low
                  stringency conditions."
BASE COUNT      28 a 34 c 43 g 59 t 1 others
ORIGIN
Query Match      78.2%; Score 17.2; DB 10; Length 165;
Best Local Similarity 86.4%; Pred. No. 7.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GCAAAGCAGCTAGCTTCAGTGTG 22
|||||
Db 42 GCAAAGCAGCTAGCTTCAGTGTG 21

RESULT 9
BH058084
LOCUS         RPCI-24-337N15-TV RPCI-24 Mus musculus genomic clone RPCI-24-337N15
DEFINITION    RPCI-24-337N15-TV RPCI-24 Mus musculus genomic clone RPCI-24-337N15
ACCESSION     BH058084
VERSION       BH058084.1 GI:14866697
KEYWORDS      GSS.
SOURCE        house mouse.

```

```

FEATURES
source
Location/Qualifiers
1. .256
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cSTB41L1"
/clone_lib="potato leaves and petioles"
/tissue_type="leaflets and petioles"
/dev_stage="8 weeks old plants"
/lab_host="SOLR"
/notes="Vector: pBlueScript SK(-); Site 1: EcoRI; site 2:
XhoI; Tissue was supplied by Dr. Fry (Cornell University).
Leaflets and petioles were isolated from 8 week old
greenhouse grown plants. The plants were watered and
fertilized freely. The tissue was immediately frozen in
liquid nitrogen."
60 a 45 c 50 g 101 t
BASE COUNT
ORIGIN
Query Match 78.2% Score 17.2; DB 12; Length 256;
Best Local Similarity 86.4%; Pred. No. 9.7e-02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

[illegible]

TITLE	JOURNAL
Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Riken Mouse ESTs (Konno, H., et al. 1999) Unpublished (1999) Contact: Yoshihide Hayashizaki. Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-5222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp/ URL: http://genome.gsc.riken.go.jp/ Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y. Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y.	



1 GCAAAGCACATAGCTTCAGTGTG 22  
 ||||| |||||  
 264 GCAAAGCAAGAACTTCAGTGTG 285

Search completed: July 8, 2003, 09:21:38  
Job time: 1119.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 242.107 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-12  
Perfect score: 22  
Sequence: 1 ggtcatcattccagatgaagag 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl: \*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	22	100.0	22	6	AX014712 Sequence
2	22	100.0	2765	6	AX014701 Sequence
3	22	100.0	174953	2	AC094732 Rattus no
4	20.4	92.7	2583	10	AF302075 Mus muscu
5	20.4	92.7	2601	10	AF157106 Mus muscu
6	20.4	92.7	2652	10	AF302076 Mus muscu
7	20.4	92.7	2694	10	AF302077 Mus muscu
8	20.4	92.7	2892	10	AF157105 Mus muscu
9	20.4	92.7	2925	6	AX033372 Sequence
10	20.4	92.7	2925	10	AF176569 Mus muscu
11	20.4	92.7	208249	2	AL607032 Mus muscu
12	19	86.4	127958	8	OSJN00063 Oryza sat
13	18.8	85.5	2076	6	AX146976 Sequence
14	18.8	85.5	2232	6	AX319864 Sequence
15	18.8	85.5	2262	6	AX146980 Sequence
16	18.8	85.5	2340	6	AX146978 Sequence
17	18.8	85.5	2340	6	AX473102 Sequence
18	18.8	85.5	2636	6	AX139743 Sequence
19	18.8	85.5	2663	6	AX139745 Sequence
20	18.8	85.5	2676	6	AX033274 Sequence
21	18.8	85.5	2714	6	AX139741 Sequence
22	18.8	85.5	2784	9	AF336981 Homo sapi
23	18.8	85.5	2850	9	AK093058 Homo sapi
24	18.8	85.5	2893	6	AX356951 Sequence
25	18.8	85.5	2893	6	AX463057 Sequence
26	18.8	85.5	2953	6	AX473100 Sequence
27	18.8	85.5	2975	6	AX356955 Sequence
28	18.8	85.5	2975	6	AX463058 Sequence
29	18.8	85.5	25807	9	AL589746 Human DNA
30	18.8	85.5	62793	2	AC131501 Lytechinu
31	18.8	85.5	144907	2	AC131453 Strongylo
32	18.8	85.5	151648	2	AC123802 Mus muscu
33	18.8	85.5	179191	2	AC102290 Mus muscu
34	18.4	83.6	78132	2	AC022794 Homo sapi
35	18.4	83.6	90150	9	AC068576 Homo sapi
36	18.4	83.6	139634	2	AC127829 Rattus no
37	18.4	83.6	159927	9	AL353590 Human DNA
38	18.4	83.6	169675	9	AC060766 Homo sapi
39	18.4	83.6	181561	9	AC015911 Homo sapi
40	18	81.8	645	6	AX380527 Sequence
41	18	81.8	134205	9	HS79664 Human DNA
42	18	81.8	172097	2	AC112880 Rattus no
43	18	81.8	189576	2	AL772183 Mus muscu
44	18	81.8	202565	2	AL772374 Mus muscu
45	17.8	80.9	10874	1	AE007242 Sinorhizo

## ALIGNMENTS

RESULT 1  
AX014712  
LOCUS AX014712 22 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 12 from Patent WO9953077.  
ACCESSION AX014712  
VERSION AX014712.1 GI:10040985  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
1 (bases 1 to 22)  
REFERENCE Schwartz,J.C., Gros,C., Ouimet,T., Rose,C., Bonhomme,M.C. and  
AUTHORS Facchinetti,P.  
TITLE Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy

JOURNAL Patent: WO 9953077-A 12 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
1. 22  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
7 a 4 c 6 g 5 t  
BASE COUNT 7 a 4 c 6 g 5 t  
ORIGIN  
Query Match 100.0%; Score 22; DB 6; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTCAATCCAGATGAAG 22  
Db 1 GGTCAATCCAGATGAAG 22  
RESULT 2  
AX014701/c  
LOCUS  
DEFINITION Sequence 1 from Patent WO9953077.  
ACCESSION AX014701  
VERSION AX014701.1 GI:10040975  
KEYWORDS  
SOURCE black rat.  
ORGANISM Rattus rattus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 2765)  
REFERENCE  
AUTHORS Schwartz, J.C., Gros, C., Oumet, T., Rose, C., Bonhomme, M.C. and  
Facchinetti, P.  
TITLE Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy  
JOURNAL Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES Location/Qualifiers  
1. 2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107. 2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSESSVGMERADNCRRRLGFVECCILLVLLTLLILMGAVTL  
GVFISIGLPLNLSLHVRHRTVVRKLDSDSKSDICTTPSCVIAAARILQNM  
OSKPCDFQYACGGLRHVHVPETNSRYSDVLDRLDEVLKGVLSQVQRHPA  
VEKATLRSQNSQVIEKRDSEPLNLDVLDIGWPVAMKWNMTGPKWELERQIAV  
LNSQNRRLVLDLFTWDDONSRRHVYIDOPTLGMPREYVFKDSHRVRAAYLQFM  
TSVATMLRDNLPCETDLVQEMAQVLHLETHLANATVQEKRDHVTALYHRMGLE  
LQERGLGFWNTLFIQNVLSVQVELLPNEEYVYIGIPLNLEEIIIDVFAQILQ  
YLWRLVLDRLGSLQKFAVDYKALYGTTEVRWRECVSYVNSMESAVGSLY  
IKRAFSDKSYSLIEKIRSVFDVLDLNMDESKKQAKRALNIREQIGPYD  
ILEDNNRHLDSEYSLTSEDYLFENGLQNLKNAQSKLRKREVDQNLWIGAAV  
NAFYSPNRLVFPAGILQPPFSDQPOLNFGGIVGVIGHEITHGFDNCRNFDKN  
GNMLDWNFSNHRPQSOQMIYQSFNSWELADONVNGFTSGENATDVHVRQA  
YKAYLQWLAEGRDRLGLNLTYAQLFQVWGSYRFEPAFQISIKTDVHVSPLK  
YRVLSGLQNLPGFSAFHCPGSPHMPNRCRIW"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 22; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTCAATCCAGATGAAG 22

Db 818 GGTCAATCCAGATGAAG 797  
RESULT 3  
AC094732  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-516, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
56 unordered pieces.  
AC094732 GI:17941511  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS Rattus norvegicus  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus  
1 (bases 1 to 174953)  
REFERENCE  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J.,  
Benton, J., Blinze, K., Blankenburg, K., Bonnin, D., Bouck, J.,  
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,  
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,  
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,  
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,  
Coyte, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,  
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,  
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,  
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,  
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,  
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,  
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,  
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,  
Hollins, B., Homs, F., Howard, S., Huber, J., Huly, S., Hume, J.,  
Joudah, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W.,  
Louiseged, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,  
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Monabat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,  
Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojokan, I., Rolfe, M.,  
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshtari, N.,  
Stinson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H.,  
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Thomas, N.,  
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.  
Direct Submission  
Unpublished  
REFERENCE 2 (bases 1 to 174953)  
AUTHORS Worley, K.C.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15624568.  
COMMENT  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
----- Project name: GBGF

```

Center clone name: CH230-516
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 152255 bases at least Q40
Consensus quality: 158448 bases at least Q30
Consensus quality: 164461 bases at least Q20
Estimated insert size: 155965; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hqsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1
13782: contig of 13782 bp in length
13783: 13882: gap of unknown length
13883: 23287: contig of 9405 bp in length
23288: 23387: gap of unknown length
23388: 28081: contig of 4694 bp in length
28082: 28181: gap of unknown length
28182: 33807: contig of 5626 bp in length
33808: 39271: contig of 5364 bp in length
39272: 39371: gap of unknown length
39372: 44270: contig of 4899 bp in length
44271: 44370: gap of unknown length
44371: 47723: contig of 3353 bp in length
47724: 47823: gap of unknown length
47824: 53427: contig of 5604 bp in length
53428: 53528: gap of unknown length
53528: 57303: contig of 3776 bp in length
57304: 57403: gap of unknown length
57404: 62118: contig of 4615 bp in length
62119: 66676: contig of 4558 bp in length
66677: 70201: contig of 3425 bp in length
70202: 70301: gap of unknown length
70302: 74082: contig of 3781 bp in length
74083: 74182: gap of unknown length
74183: 78306: contig of 4124 bp in length
78307: 78406: gap of unknown length
78407: 81222: contig of 2816 bp in length
81223: 81322: gap of unknown length
81323: 84552: contig of 3230 bp in length
84553: 84652: gap of unknown length
84653: 88839: contig of 4187 bp in length
88840: 89399: gap of unknown length
89400: 92256: contig of 3317 bp in length
92257: 92356: gap of unknown length
92357: 94995: contig of 2639 bp in length
94996: 95095: gap of unknown length
95096: 98465: contig of 3370 bp in length
98466: 98565: gap of unknown length
98566: 101632: contig of 3067 bp in length
101633: 101732: gap of unknown length
101733: 104939: contig of 3207 bp in length
104940: 105039: gap of unknown length
105040: 109248: contig of 4209 bp in length
109249: 109348: gap of unknown length
109349: 112435: contig of 3087 bp in length
112436: 112535: gap of unknown length
112536: 114676: contig of 2141 bp in length
114677: 114776: gap of unknown length
114777: 117567: contig of 2791 bp in length
117568: 117667: gap of unknown length

```

```

* 117668 119209: contig of 1542 bp in length
* 119210 119309: gap of unknown length
* 119310 121797: contig of 2488 bp in length
* 121798 121897: gap of unknown length
* 121898 124293: contig of 2396 bp in length
* 124294 124393: gap of unknown length
* 124394 124428: contig of 2035 bp in length
* 124429 125228: gap of unknown length
* 125229 125525: contig of 2997 bp in length
* 125526 129625: gap of unknown length
* 129626 132334: contig of 2709 bp in length
* 132335 132434: gap of unknown length
* 132435 132774: contig of 2840 bp in length
* 132775 133074: gap of unknown length
* 133075 138074: contig of 2700 bp in length
* 138075 138174: gap of unknown length
* 138175 139985: contig of 1811 bp in length
* 139986 140085: gap of unknown length
* 140086 142273: contig of 2188 bp in length
* 142274 142373: gap of unknown length
* 142374 143598: contig of 1225 bp in length
* 143599 143698: gap of unknown length
* 143700 14534: contig of 1736 bp in length
* 14535 145534: gap of unknown length
* 145535 146985: contig of 1451 bp in length
* 146986 147085: gap of unknown length
* 147086 148199: contig of 1014 bp in length
* 148200 150915: contig of 2716 bp in length
* 150916 151015: gap of unknown length
* 151016 152501: contig of 1486 bp in length
* 152502 152602: contig of 1409 bp in length
* 154011 154110: gap of unknown length
* 154111 155758: contig of 1648 bp in length
* 155759 155858: gap of unknown length
* 155859 157622: contig of 1764 bp in length
* 157623 157722: gap of unknown length
* 157723 159428: contig of 1706 bp in length
* 159429 159528: gap of unknown length
* 159529 161209: contig of 1681 bp in length
* 161210 161309: gap of unknown length
* 161310 163413: contig of 2104 bp in length
* 163414 163513: gap of unknown length
* 163514 164702: contig of 1189 bp in length
* 164703 164802: gap of unknown length
* 164803 165998: contig of 1196 bp in length
* 165999 166099: gap of unknown length
* 166100 167413: contig of 1314 bp in length
* 167414 167512: gap of unknown length
* 167513 169231: contig of 1719 bp in length
* 169232 169331: gap of unknown length
* 169332 170534: contig of 1203 bp in length
* 170535 170634: gap of unknown length
* 170635 172048: contig of 1413 bp in length
* 172049 172148: gap of unknown length
* 172149 173509: contig of 1362 bp in length
* 173510 173609: gap of unknown length
* 173610 174953: contig of 1344 bp in length.

```

## FEATURES

```

Query Match      100.0%; Score 22; DB 2; Length 174953;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  GGTCATCATTCAGATGAAGAG 22
          |||||

```

```

Db      60879  GGTCATCATTCAGATGAAGAG 60900

```

## RESULT 4

```

AF302075/c

```

```

LOCUS

```

```

AF302075

```

```

2583 bp

```

```

mrna

```

```

linear

```

```

ROD 11-JUN-2001

```



```
/codon_start=1
/product="neprilysin-like peptidase gamma"
/protein_id="AAG18448.1"
/db_xref="GI:10505364"
/translatio="MYERAGWCRRKSPGFVEYGLMVLILLILLAGVTLGTVFYFYSIALRD
SLSKDICTPTSCVTAAARILENMQDSRNPENFYQACGGWLRHVHIVETNSRVSYVF
DIIRDLELVLGVDLSTQSQRPAVEKARTLYRSCMNOSVIIEKRDESEPLLVSLLKVMG
GWPMAMDKNWTMTGLSKWELOALPVANSOFNNRWRLIDLFIWNDDQNSSRHVYIQDPT
LGMPSRETYFOEDNNHKVKAYLEFTSVATMLRKDONLSKSASAMRYEAMAEVYLET
HLANATYPOEKHRDHVALYHRMDLMELQSRFGIKDRVSLCSPGCPTHSDVOAGALET
NPASDSRVGLKGFWNTLFIONVLSSEVELFPDDEVVVYGIPYLLENLEDIIDTSYA
RTMONYLVWLVIDRGISLSORFKEARVDYRKALYGTTEVERWRCEVCNVSNMBESA
VGSLYIKRAFSDKSTSRYRELIEKIRSFVEDNLDELNMWDERSKKKAQBKANRIEQI
GPFDILEDNNKHDEESYSLTFIEDLIYFENGQLNKNAQRSKLKLRBKVDONLMIWI
GAADVNFYPNRNDIOFPSFPGILOPPFSKDPQSQLNGFGIGIMGVITGEITHGHGDNRG
NDFDKINGMLDMWSNSSRAHFQQOCSMICYYGNFSELADNONVANGFSYFLGSNIADNG
GYROYAKYLRWLADGGCKDORPLGLNLVAOLFFNYAOVWGCGSYRPPEFAVQSIIKTID
HSPLYKRVLSGLNQLPDGFEAPFCPRGPSMPMHMKCRHW"
```

BASE COUNT            686 a      700 c     766 g       542 t

Query Match                  92.7%; Score 20.4; DB 10; Length 2694;  
Best Local Similarity      95.5%; Pred. No. 8.8;

Matches	21;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

OY        1 GGTCTCATTCAGATGAAGAG 22  
||||| ||||| ||||| ||||| |||||  
Db        637 GGTCGTCA TTCAGATGAAG 616

```
RESULT 8
AF157105/c
LOCUS AF157105            2892 bp      mRNA           linear       ROD 25-NOV-1999
DEFINITION Mus musculus soluble secreted endopeptidase mRNA, complete cds.
ACCESSION AF157105
VERSION   AF157105.1    GI:6467398
KEYWORDS SOURCE
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus..
          1 (bases 1 to 2892)
Ikeda,K., Emoto,N., Raharjo,S.B., Nuhantari,Y., Saiki,K.,
Yokoyama,M. and Matsuo,M.
Molecular identification and characterization of novel
membrane-bound metalloprotease, the soluble secreted peptides
hydrolyzes a variety of vasoactive peptides
J. Biol. Chem. 274 (45), 32469-32477 (1999)
20011457
10542292     2 (bases 1 to 2892)
Ikeda,K., Emoto,N. and Matsuo,M.
Direct Submission
Submitted (08-JUN-1999) International Center for Medical Research,
 Kobe University School of Medicine, 7-5-1 Kusunoki, Chuoh, Kobe
        6500017, Japan
FEATURES             Location/Qualifiers
     source            .
         1 ..2892
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             281..2578
                 /note="SEP; metallopeptidase"
                     /codon_start=1
                         /product="soluble secreted endopeptidase" ,
                             /protein_id="AAF13152.1"
                                 /db_xref="GI:6467399"
                                     CDS
                                         ORIGIN
```

BASE COUNT 717 a 774 c 827 g 574 t  
 ORIGIN  
 Query Match 92.7%; Score 20.4; DB 10; Length 2892;  
 Best Local Similarity 95.5%; Pred. No. 8.8;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGTATCATTCAGATGAAG 22  
 Db 962 GGTGTCATTCAGATGAAG 941

RESULT 9  
 AX033272/c  
 LOCUS AX033272 2925 bp DNA linear PAT 21-SEP-2000  
 DEFINITION Sequence 12 from Patent WO0047750.  
 ACCESSION AX033272  
 VERSION AX033272.1 GI:10280087  
 KEYWORDS house mouse.  
 SOURCE Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 AUTHORS 1 (bases 1 to 2925)  
 TITLE Boileau, G. and Desrosiers, L.  
 JOURNAL New metalloproteases of the neprilysin family  
 PATENT: WO 0047750-A 12 17-AUG-2000;  
 BOILEAU GUY (CA); DESGROSEILLERS LUC (CA); UNIVERSITE DE MONTREAL (CA)

FEATURES  
 Source Location/Qualifiers  
 1..2925  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 332..2629  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC09977.1"  
 /db\_xref="GI:10280088"

/translation="MVERAGWCRKSPGVEYGLMVLILLGALITVLGVFYSIGKQL  
 PLLTSLHFWSDERTVVKRALRDSLSKSDICTPSCVIAAARILENMDQSRNPNCFY  
 QYACGWLRRHVIPETNSRYSVFDILRDELEVILKGVLESDTSOHRPAVEKAKTLVRS  
 CMNQSVIEKRDESEPLLSVLKVGWGPVAMDKWNETMGLKWELEKRLAVLNSQFNRRVL  
 IDLFITNDQNSRHVIYIDOPTLGMPSREYFQEDNNHKVKAYLEFMTSVATMLRK  
 DQNLKESAMVREMAEVELETHLANATVPQEKHDVTALYHRMDLMELOERGLKG  
 FNWTLFIQNLVSSVEVEFPDEEVVYIGIYLENEDIISYARTMNYLWRLVLD  
 RIGLSQRFKEARVDYRKALYGTVEEVRWRECVSYVNMESAVGSLYIKRAFSDKS  
 KSTVRELIEKRSVFDNLDELNMWDEESKKAQKAMNIREQIGYDYLIEDNNKHL  
 DEYSSLFYEDLYFENGLQNLKNAQSRSLKLRKVDQNLATIGAAVNAVYSPNRR  
 QIVFPAGILQPPFSKQDQPSLNFQGGIGWIGHEITHGFDNNGRDNKNGMLDWSN  
 FSARHFQOQSOCHMIYQYGNFSEWELADNQNNGVSTLGENIADNGGVRQAYKAYLRWLA  
 DGGKQRLPGLNLTYAQLFFINAYQVWCGSYRPEFAVQSIKTDVHSPKRYRVLGSLQ  
 LPGFEAFHCPRGSPMHPMKCRW"

BASE COUNT 710 a 797 c 836 g 582 t  
 ORIGIN

Query Match 92.7%; Score 20.4; DB 6; Length 2925;  
 Best Local Similarity 95.5%; Pred. No. 8.8;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGTATCATTCAGATGAAG 22  
 Db 1013 GGTGTCATTCAGATGAAG 992

RESULT 10  
 AF176569/c

LOCUS AF176569 2925 bp mRNA linear ROD 11-MAY-2000  
 DEFINITION Mus musculus neprilysin-like metalloproteinase 1 (N1L) mRNA,  
 complete cds.  
 ACCESSION AF176569  
 VERSION AF176569.1 GI:7769082  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 2925)  
 AUTHORS Ghaddar, G., Ruchon, A. F., Carpentier, M., Marcinkiewicz, M.,  
 Seidah, N. G., Crine, P., Desrosiers, L. and Boileau, G.  
 TITLE Molecular cloning and biochemical characterization of a new mouse  
 testis soluble-zinc-metalloproteinase of the neprilysin family  
 JOURNAL Biochem. J. 347 (Pt 2), 419-429 (2000)  
 MEDLINE 21060448  
 PUBMED 10749671

REFERENCE 2 (bases 1 to 2925)  
 AUTHORS Ghaddar, G., Ruchon, A. F., Desrosiers, L. and Boileau, G.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-AUG-1999) Biochemistry, University of Montreal, 2900  
 Edouard Montpetit, Montreal, Que H3T 1J4, Canada  
 FEATURES Location/Qualifiers  
 1..2925  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /tissue\_type="testis"  
 1..2925  
 /gene="N1L"  
 332..2629  
 /gene="N1L"  
 /note="N1L; metallo-endopeptidase"  
 /codon\_start=1  
 /product="neprilysin-like metalloproteinase 1"  
 /protein\_id="AAF69247.1"  
 /db\_xref="GI:7769083"

/translation="MVERAGWCRKSPGVEYGLMVLILLGALITVLGVFYSIGKQL  
 PLLTSLHFWSDERTVVKRALRDSLSKSDICTPSCVIAAARILENMDQSRNPNCFY  
 QYACGWLRRHVIPETNSRYSVFDILRDELEVILKGVLESDTSOHRPAVEKAKTLVRS  
 CMNQSVIEKRDESEPLLSVLKVGWGPVAMDKWNETMGLKWELEKRLAVLNSQFNRRVL  
 IDLFITNDQNSRHVIYIDOPTLGMPSREYFQEDNNHKVKAYLEFMTSVATMLRK  
 DQNLKESAMVREMAEVELETHLANATVPQEKHDVTALYHRMDLMELOERGLKG  
 FNWTLFIQNLVSSVEVEFPDEEVVYIGIYLENEDIISYARTMNYLWRLVLD  
 RIGLSQRFKEARVDYRKALYGTVEEVRWRECVSYVNMESAVGSLYIKRAFSDKS  
 KSTVRELIEKRSVFDNLDELNMWDEESKKAQKAMNIREQIGYDYLIEDNNKHL  
 DEYSSLFYEDLYFENGLQNLKNAQSRSLKLRKVDQNLATIGAAVNAVYSPNRR  
 QIVFPAGILQPPFSKQDQPSLNFQGGIGWIGHEITHGFDNNGRDNKNGMLDWSN  
 FSARHFQOQSOCHMIYQYGNFSEWELADNQNNGVSTLGENIADNGGVRQAYKAYLRWLA  
 DGGKQRLPGLNLTYAQLFFINAYQVWCGSYRPEFAVQSIKTDVHSPKRYRVLGSLQ  
 LPGFEAFHCPRGSPMHPMKCRW"

BASE COUNT 710 a 797 c 836 g 582 t  
 ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 2925;  
 Best Local Similarity 95.5%; Pred. No. 8.8;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGTATCATTCAGATGAAG 22  
 Db 1013 GGTGTCATTCAGATGAAG 992

RESULT 11  
 AL607032/c  
 LOCUS AL607032 208249 bp DNA linear HTG 24-JUL-2002  
 DEFINITION Mus musculus chromosome 4 clone RP23-15L19, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 2 unordered pieces.  
 ACCESSION AL607032  
 VERSION AL607032.15 GI:21955491  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

BASE COUNT 710 a 797 c 836 g 582 t  
 ORIGIN

Query Match 92.7%; Score 20.4; DB 10; Length 2925;  
 Best Local Similarity 95.5%; Pred. No. 8.8;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGTATCATTCAGATGAAG 22  
 Db 1013 GGTGTCATTCAGATGAAG 992

RESULT 11  
 AL607032/c  
 LOCUS AL607032 208249 bp DNA linear HTG 24-JUL-2002  
 DEFINITION Mus musculus chromosome 4 clone RP23-15L19, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 2 unordered pieces.  
 ACCESSION AL607032  
 VERSION AL607032.15 GI:21955491  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus  
 ORGANISM Mus musculus

BASE COUNT 710 a 797 c 836 g 582 t  
 ORIGIN

Query Match 92.7%; Score 20.4; DB 6; Length 2925;  
 Best Local Similarity 95.5%; Pred. No. 8.8;  
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGTATCATTCAGATGAAG 22  
 Db 1013 GGTGTCATTCAGATGAAG 992

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 208249)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Direct Submission  
Submitted (23-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jul 25, 2002 this sequence version replaced gi:21627906.

COMMENT

----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: humquerry@sanger.ac.uk  
----- Project Information  
Center project name: BM15L19

----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator; 11% of reads  
Chemistry: Dye-terminator; Big Dye; 88% of reads  
Consensus quality: 207834 bases at least Q40  
Consensus quality: 207956 bases at least Q30  
Consensus quality: 208060 bases at least Q20  
Insert size: 208149; sum-of-contigs  
Insert size: 219051; 2.3% error; agarose-ff  
Quality coverage: 12.41x in Q20 bases; sum-of-contigs Quality  
coverage: 12.08x in Q20 bases; agarose-ff

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved

\* 1 137303: contig of 137303 bp in length  
\* 137304 137403: gap of 100 bp  
\* 137404 208249: contig of 70846 bp in length.

FEATURES

source

1. .208249  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="4"  
/clone="RP23-15L19"  
/clone\_lib="RPC1-23"  
misc\_feature  
1. .137303  
/note="assembly\_fragment:04071"  
misc\_feature  
137404..208249  
/note="assembly\_fragment:01838  
clone\_end:77  
vector\_side:right"  
vector 53647 c 51328 g 49076 t 100 others

BASE COUNT 54098 a 53647 c 51328 g 49076 t 100 others

ORIGIN

Query Match 92.7%; Score 20.4; DB 2; Length 208249;  
Best Local Similarity 95.5%; Pred. No. 9.3;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAGAG 22  
|||||

Db 173629 GGTCTGTCATTCAGATGAGAG 173608

RESULT 12

OSJN00063

LOCUS 127958 bp DNA linear PLN 12-JUL-2002  
DEFINITION Oriza sativa genomic DNA, chromosome 4, BAC clone: OSJNBB0086G13,  
complete sequence.

ACCESSION AL506706

VERSION AL506706.2 GI:21741205

KEYWORDS HTG.

SOURCE  
ORGANISM

Oriza sativa  
Oriza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE  
AUTHORS

Fu, G., Wang, S. Y., Ren, S. X., Lv, G., Lin, W., Gu, W. Q., Zhu, G. F.,  
Tu, Y. F., Jia, J., Yin, H. F., Zhang, Y., Cai, Z., Chen, J., Kang, H.,  
Chen, X. Y., Shao, C. Y., Sun, Y., Hu, Q. P., Zhang, X. L., Zhang, W.,  
Wang, L. J., Ding, C. W., Sheng, H. H., Gu, J. L., Chen, S. T., Ni, L.,  
Zhu, F. H., Han, B., Feng, Q., Huang, Y. C., Li, Y., Zhu, J. J., Zhao, Q.,  
Hu, X., Liu, Y. L., Mu, J., Yu, Z., Chen, L., Fan, D. L., Weng, Q. J.,  
Zhang, L., Lu, Y. Q., Yu, S. L., Liu, X. H., Lu, T. T., Zhang, Y. J., Lu, Y.,  
Li, C., Li, T., Zhang, Y., Hu, H., Jia, P. X., Qian, Y. M., Ying, K.,  
Zhou, B., Chen, Z. H., Hao, P., Zhang, L., Wu, M., Zhang, R. Q., Guan, J. P.  
and Hong, G. F.

TITLE  
JOURNAL

Direct Submission  
Submitted (27-JUN-2002) Han Bin, National Center for Gene Research,  
Chinese Academy of sciences, 500# Cao-Bao Road, Shanghai 200233,  
CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:  
bhan@ncgr.ac.cn

REMARK

Oriza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC  
clone: OSJNBB0086G13.

COMMENT

On Jul 12, 2002 this sequence version replaced gi:15620782.

Web site: <http://www.ncgr.ac.cn>

----- Summary Statistics

Assembly program: phrap

NOTE: This is a complete sequence.

Genes were identified by a combination of several methods: Gene  
Prediction programs including Fgenesh (<http://www.softberry.com/>),  
genscan (<http://CCR-081.mit.edu/GENSCAN.html>), GenemarkHMM  
(<http://genemark.biology.gatech.edu/Genemark/>), tRNAscan-SE (Sean  
Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>), searches of the  
complete sequence against NCBI none redundant protein database (nr)  
(<ftp://ncbi.nlm.nih.gov/blast/db>) and the EST database at NCGR.

FEATURES

source

1. .127958  
/organism="Oriza sativa"  
/variety="Nipponbare"  
/sub\_species="japonica"  
/db\_xref="taxon:4530"  
/chromosome="4"  
/clone="OSJNBB0086G13"  
/clone\_lib="CUGI-OSJNBB"  
complement(184..2118)  
/gene="OSJNBB0086G13.1"  
complement(join(184..261,357..512,992..1073,1554..1893,  
2004..2118))  
/gene="OSJNBB0086G13.1"  
/codon\_start=1  
/protein\_id="CAD41017.1"  
/db\_xref="GI:21741206"  
/translation="MDAVMSAADGAGSREDPPPVVLVSAGASHSALLAGNVL  
SWGDEGQILGHGADRLVPLVSGFDAAAGPITSVICGADHTTAYSEDDQVYSWG  
WGDFGRHGNSDVFPPQPVKQGIKIQIACGSHCLAVTWAGEVQSGRNGQ  
LGLGTEDSLLPKQIQSGFVGCVKMAIAGAEHTAAVTEDGLYGVGWGRYGNLGLGR  
NDRLPKSVSSVENMFLKIPSLYATGLYCQSVTISLLS"  
21732..22637  
/gene="OSJNBB0086G13.2"  
-join(21732..21940,22049..22637)  
/gene="OSJNBB0086G13.2"  
/codon\_start=1  
/protein\_id="CAD41018.1"  
/db\_xref="GI:21741207"  
/translation="MQLYHPYSLDSOKVRMALEEKIDYTSYHVNPLTGKMNVAFF  
RMNPSAKLPVFGAHVIYRAFDTIQIDRLSLHLSGEIVPVTVEYQWQKVSNNP  
KMTLTHTPKYRTFVSKFIRRVLIARMAEPDLASMHAKIRAYETEDKLKPDIM  
KQSEELSKLDDVEQLNNGKYLAGDEFSADSVFIPILARITLLDDEYINCRP  
LLEYLYLKORPSYKVAIGFPGGWKKYRILFKTSFFLCVTLFRKY"  
complement(42703..47094)  
/gene="OSJNBB0086G13.3"  
complement(join(42703..43413,44264..44569,44768..44959,

gene

CDS

45284..45403,46511..46693,47014..47094))

/gene="OSJNBb0086G13.3"  
/codon\_start=1  
/db\_xref="GI:21741208"  
/translation="MAGEVAIDSGEKLNELGYKQELREMTLFTKLALSFMTWLTFT  
GTPLYGSLRTYPASLVGWGVVQVFFTFVGFMAEICSFFPTGSLYFWAAHLG  
PWPLASCCAWLEAGLIGAGTQAGSVOFIASILLCTGKNGGGLYLPWFLF  
LMYIGLTFWALNFALEVIADFLISMMWVIGTVIVLPLVAKTQIPASTVFT  
HFEAPEATGIRSSAVARTISLLVSQSLYDAAHLTEETKGDKNKPIALSSIG  
LITVGMAYILALTESIQDFSLFDPSENFAGTFFPAQILFPAFGRGSSAGATALL  
FVMSGFFLGUSITTSARVYALSROGVFPLSVWRVRRHPRFANAWLCAAC  
ALLGPIILWINVITATSIATIGWVGTVAFIFARVMREDFSPGPYLLRASRPV  
RIDAPLWCTCTVFLFTPTAPISAGNFYAPVALGACGLIGLWVLDARWRFGPV  
RNIDPLKGGDGGVHNNKVV"  
/complement(51054..55507)  
/gene="OSJNBb0086G13.4"  
/complement(join(51054..51064,52665..52692,52912..52939,  
53215..53309,54838..54993,55148..55507))  
/gene="OSJNBb0086G13.4"  
/codon\_start=1  
/protein\_id="CAD41020.1"  
/db\_xref="GI:21741209"  
/translation="MPLVVGCPRRGPPRPAAGRAORPPPHSVAPKRRSSSSSTTTP  
LITTRAAASALLAALPPFPASSPOLPVASAKEDDAEAREGELELELELER  
YDRDEGFTLKPVPVKYKAGATLFOQEGKGSNNIGIVVNPVRLSLTLTEFGTPQF  
VAERLQAEKKKIQDMVAAHIRKFLRLKVHTSIDHPKYSALITRSSVADQDYQVE  
SSRSDAI"  
57890..62229  
/gene="OSJNBb0086G13.5"  
/join(57890..58066,58171..58329,59427..59586,59633..59733,  
59852..59939,60259..60326,60423..60509,61006..61080,  
61169..61246,61374..61478,62143..62229)  
/gene="OSJNBb0086G13.5"  
/codon\_start=1  
/protein\_id="CAD41021.1"  
/db\_xref="GI:21741210"  
/translation="MAORLAALHAAPSAFSTSSASGRPAARSSTTALLPVALPR  
ASATLRAAPSLRLLPQEAAGSRVCMASASAASAAVASGAELKAAREDIRL  
LKTTHCHPLVLRLGHDSTVDKNIKWPORGANGSLREFVELKHGANAGNITSRF  
RPLVWLYNALKLVPIDKYPNISYADLQLASATATAEAGCPKIPMTYGRIDVTGP  
ECCPEGLPDLGPADADHLKRYEIRMGLDKEIIVLSAGTPIRGSGWKPE  
TKYKNGPAGGGSTAEWLFDNSYFKEIKERKQDLVLPLTDAALFEDPTFFKYA  
EKYAEQDAFFTADYAGAHAKLSNLGAKFNPPELYQPDAPPTTGAADLQTAHTPOGPG  
HE"

gene

CDS

gene

CDS

gene

CDS

64565..76078  
/gene="OSJNBb0086G13.6"  
/join(64565..64953,65055..65152,65923..66012,66531..66583,  
66919..66972,67150..67209,67463..67541,67838..67905,  
68036..68095,68204..68308,68708..68812,68945..69052,  
69120..69266,69481..69663,70037..70111,70435..70605,  
70846..70950,72016..72252,72720..72860,72991..73116,  
73372..73422,73521..73568,73645..73881,73985..74662,  
74852..74941,76055..76078)  
/gene="OSJNBb0086G13.6"  
/codon\_start=1  
/protein\_id="CAD41022.1"  
/db\_xref="GI:21741211"  
/translation="MSSSRPGRASIPFRSRRTSAAGGAGVAAAHPPTARTSGG  
RPSTPSSSSAGGRPTTPSSSAGGRPTTSAFAFARSTTPSSGRPTTPSSASRAA  
GRAPPVAAANAKENITVTRPLSPREINKGDEVANVANGDMVRNENYSIAY  
AFDKVGPATTHRDYIDAAOHNRREDAGTYVEGKEEVVLSAPAHLSIASEEH  
RVSYLEIYNEVINDLDPAGONLRREDAGTYVEGKEEVVLSAPAHLSIASEEH  
RVSGNNENLYSSRSHTIFTLTIESPSGDEGEVKLSQLNLIDLAGSESSKETTGG  
LRRKGSYINKSLLTIVIAKLDGKATHIPRIQSKLRLQLQSSLSGHGRSLICTV  
TPASNSETHNLKFAHRSHIEIKASQNKIIDKSLIKKYQKTCILKEELQLLR  
GMNGYIPTPDQDLVLSKLEAGQVKLQSLRELEEEAKAALMGRQLRKLILVS  
TKSSITSSVSGKASLRRRHSFGEDEKPTIYVMFWLLTSELLDMRELAYLDPDRK  
REYSMEDDVSLDSFEVSGKLDNNPDESLEFRNRRRGLHFKLKNHISNIRK  
LDKYSKCELETVVGVFVIVATHVKNQSLQSLSTVDSSTASGSPFSRSQQKH  
PLLDLUDKRRKSMRKGDPALDTPFGQTQAGDLFSAASRAHRHLKSHIRYLEQ  
LOEQVLMAGEVALCTSSKRLSEQAANNPDQSIOEQTEKLNIDEKKSHIRYLEQ  
RMAQSLETTEDPAINTSQFSKLSTOLSEKTFELTMSADNRILQLOLQAKVSENA

gene

CDS

gene

CDS

gene

CDS

ELVETVAQRQRIDNLLKAKNEDNVASOSSPSTSSNPRDLANEVASHKMPST  
TEDHTESPLKSVLLQAAREIENKDKLAEKDGLEIHSQKLAESSYAKELAA  
AVELKLAEEVTRLSYENAKLNADLAAKDDQTRSSIQSDTKRQDQENGIVAELOKEL  
VASCOREAVLEDTLSQARRESLELLAVIEDAKCHEHDLLENELAMNMLVAELKENSO  
EDLFQFKATONGYHSSKSDTGRMKSMEASDENWDGVSYSYEEAKAANYVORRCK  
ELGIVSRUKGEDRGDKVLEELQNFHVEALSKICQEAQVDVNEKD"

gene

CDS

/complement(87600..89253)  
/gene="OSJNBb0086G13.7"  
/complement(join(87600..87950,88030..89253))  
/gene="OSJNBb0086G13.7"  
/codon\_start=1  
/protein\_id="CAD41023.1"  
/db\_xref="GI:21741212"  
/translation="MNVVAIGLVLSLAAAGVWSPAPTPPPSTAGEHVVRREGRRVVI  
VYERHPLYPGDVKVTHVLPDSDLDGEGRLTKARGAVSDAARDVADAAGEGKKE  
MSDAKESATGKVFAGVKCRCKDLRCCGAAREAEKAKDKASAGEHGAEEAAGEALSH  
AKEIADKVFDAASKVKETAVGAKDKVSKAKAKERASHVQHGAETVRNAKDKVSH  
AARHAKESRERAMDAKRVSDVAERAEQCTEDDAAGRAAKAAGAAEAVKAKAGEAAS  
NLSDTARRARDVADAAHLGGPREAARTATVMHLGLFRAAAYGASLWTVTSVV  
LAAALPQQLAMVQSKLFPMYFRAVAYGVGLALAAHLGFRERSLAARQASVNLAAAL  
ALVANMLLELPKATKVMFERMKVEGGRMDADIVDPTVTVAAGNTATTITVPT  
AARAKPAGNNMAAKSAAAAAPVDEQTSKSRVATLNQRLKKLNGYSLSLNCVLCMLALT  
WHYHLARLQIASVC"

gene

CDS

89836..93186  
/gene="OSJNBb0086G13.8"  
/join(89836..90009,90409..90500,90820..91069,91124..91339,  
91425..91500,91886..92038,92176..92368,92361..92521,  
92629..93186)  
/gene="OSJNBb0086G13.8"  
/codon\_start=1  
/protein\_id="CAD41024.1"  
/db\_xref="GI:21741213"  
/translation="MEMEDIVLGPAGIAGGAAPCLRLPLAAVAVKPKRPSRVA  
OTRPOEARIPGTSDEYMSQLSAFYAITPEPECADLWLINTCTVKKPSQSAMTT  
LISKCSANKPLVWAGCVQGSRLDELEGISVIGVQIDRVVEVEETLKGHEVRL  
SRKTLPSLDLPKYSANVMCYIYFTDRCNLISLFTMDHSYFQVRKKNFIEILPINV  
CGICACTCKTHARGHLGSIYTESLDVRKIVVSEGVREINLSSEDTGAYFAPGPD  
LPHNTSQAMNREYTVSEFRVVDVTLCELVPGQIATDIICGPPGDEDFSQTNVLV  
KOYLPFQVHISQFYPRPGTPAARMKKVPSEVKKRSRELTSVFSESPYOGMEGKVER  
IWTELATDGHVHLGHTKGYIOVLVIAPDMLGTSDVKITSVGRKSVFGEVGVV  
AKETQHNSHELOEYRPSQVEATCCGTDSCACCTSDAAQOCCPSESENSTROS  
CDGATHQAVQSKLRVNRVEGAVKSSSESOTSCGCGEIQRLNVAARRPDDVTILMGL  
AVSFATTVTLLIALLAYKISSTPSY"  
99664..101258  
/gene="OSJNBb0086G13.9"  
/join(99664..99827,100088..101258)  
/gene="OSJNBb0086G13.9"  
/codon\_start=1

gene

CDS

Query Match

Best Local Similarity 86.4%; Score 19; DB 8; Length 127958;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTCATCATTCACAGATGAAG 20

|||||

Db 61943 GTCATCATTCACAGATGAAG 61961

RESULT 13

AX146976/c

LOCUS

Sequence 1 from Patent WO0136610. 2076 bp DNA linear PAT 08-JUN-2001

DEFINITION

AX146976

VERSION

AX146976.1 GI:14346247

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 2076)

Deleersnijder W., Wieggers, R. and Weske, M.

Human enzymes of the metalloprotease family

Patent: WO 0136610-A 1 25-MAY-2001;

Solvey Pharmaceuticals B.V. (NL)  
Location/Qualifiers  
1. .2076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1. .2076  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC41159.1"  
/db\_xref="GI:14346248"  
/translation="CTTPGCVIAAARILQNMDDTPEDDYQYFCAGGMLRRHVIPET  
NSRYSIFDLRDEVLIVKALVANSYAKORPAVEKATLYRSCMNQSVLEKRGSQLP  
DILEVVGVPVMDRNETFVGLWELEERQLALMNSQFNRYLIDLFIWDDQNSSRHI  
IYIDPTLGMPSREYTPGNSRKYREAVLQFVSVATLLREDAANLPDSCLVQEDMM  
QVLEETQLAKATVPOEERHDVIALYHRMGLELQSQGLKGFNWTFLTQTVLSSVKI  
KLDPDEVVYGYIPYQLNLENIIDTYSARTIQNYLVRLVLDIRIGLSORFKDTRVY  
KRALFTWVEEVRWCQYVNSNMENAVGSLYVREAFPGDKSMYRELIDKVRTVY  
ETLDELGMDEESKKAQKAMKSIREQIGHPDYILIEEMNRRLDEEYSNLNFSEDLYK  
NSLQNLKVGAAQSLRLKREKVDPLMALIGAAVVAEYSPNRNQLVPPAGILOPPFEK  
EPOALNFGGIGNGVIGHETHGDFDDNGRFDKNGNMDDWNSNESTQHFREQSCMILY  
YGNWDLADEQVNGVITLGENIADNGVQRAYKAYLKWMEGGKQQLPLGLDLTHE  
QUFFINAGQVGSYRPEFAIOGSIKTDVHSPLYRVLGSLQNLAAFPDTHFCARGTTP  
HPKERCRVW"

BASE COUNT 493 a 578 c 636 g 369 t

ORIGIN

Query Match 85.5%; Score 18.8; DB 6; Length 2076;  
Best Local Similarity 90.9%; Pred. No. 64;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGTCATCATCCAGATGAAG 22  
||||| |||||||||  
Db 460 GGTCGTCGTCCAGATGAAG 439

RESULT 14  
AX319864/c  
LOCUS  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1  
PLOWMAN,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and

1  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
BASE COUNT  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
20; Conservative  
0; Gaps  
0;  
Qy  
1 GGTCATCATCCAGATCAAGAG 22  
|||||  
619 GGTCGTCGTTCCAGATCAAGAG 598  
Db  
RESULT 15  
AX146980/C  
LOCUS  
DEFINITION  
Accession  
AX146980  
Sequence 5 from Patent WO0136610.  
2262 bp  
DNA  
linear  
PAT 08-JUN-2001

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 30.3707 Seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-12

Perfect score: 22

Sequence: 1 ggcatcattccagatgaag 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.2	78.2	1919	1	US-08-530-492-106
C 2	17.2	78.2	1919	4	US-08-905-517-106
C 3	17.2	78.2	1975	4	US-09-052-089A-8
C 4	17.2	78.2	4072	4	US-09-770-315-4
C 5	17.2	78.2	4680	1	US-08-254-358-1
C 6	17.2	78.2	4680	1	US-08-475-391-1
C 7	17.2	78.2	4680	2	US-08-709-609-1
C 8	17.2	78.2	4680	5	PCT-US95-07178-1
C 9	17.2	78.2	7557	4	US-09-770-315-3
C 10	17.2	78.2	8698	4	US-09-770-315-2
C 11	17.2	77.3	5769	1	US-08-652-971-1
C 12	17.2	77.3	5769	2	US-08-991-258A-1
C 13	17.2	77.3	5769	2	US-08-769-399-1
C 14	17.2	77.3	5769	3	US-08-991-953A-1
C 15	16.2	73.6	11748	1	US-08-611-107-30
C 16	16.2	73.6	70000	4	US-09-851-896-3
C 17	15.8	71.8	387	4	US-09-280-116-125
C 18	15.8	71.8	7521	4	US-09-004-838-116
C 19	15.6	70.9	220	4	US-09-263-933-22
C 20	15.6	70.9	720	6	Patent No. 5310729
C 21	15.6	70.9	1467	4	US-09-330-317B-17
C 22	15.6	70.9	1467	4	US-09-808-589A-17
C 23	15.6	70.9	1956	3	US-08-867-352-20
C 24	15.6	70.9	2088	4	US-09-513-057C-4
C 25	15.6	70.9	2518	4	US-09-513-057C-1
C 26	15.6	70.9	2606	4	US-09-234-827B-3
C 27	15.6	70.9	4221	4	US-09-513-057C-3

C 28	15.6	70.9	4221	4	US-09-513-057C-34
C 29	15.6	70.9	5011	1	US-08-141-893-1
C 30	15.6	70.9	5011	1	US-08-463-092B-1
C 31	15.6	70.9	5011	1	US-08-463-092B-3
C 32	15.6	70.9	5011	2	US-08-462-109A-1
C 33	15.6	70.9	5011	2	US-08-462-109A-3
C 34	15.6	70.9	5011	2	US-08-460-907B-1
C 35	15.6	70.9	5011	2	US-08-460-907B-3
C 36	15.6	70.9	5011	3	US-08-463-179A-1
C 37	15.6	70.9	5011	3	US-08-463-179A-3
C 38	15.6	70.9	5011	3	US-08-461-384B-1
C 39	15.6	70.9	5011	3	US-08-461-384B-3
C 40	15.6	70.9	5011	3	US-08-407-207A-1
C 41	15.6	70.9	7076	4	US-09-837-863-20
C 42	15.6	70.9	7076	4	US-09-837-863-21
C 43	15.6	70.9	7092	4	US-09-837-863-19
C 44	15.6	70.9	7092	4	US-09-837-863-22
C 45	15.6	70.9	7573	4	US-09-837-863-27

ALIGNMENTS

RESULT 1  
US-08-530-492-106/C  
; Sequence 106, Application US/08530492  
; Patent No. 5689052  
; GENERAL INFORMATION:  
; APPLICANT: Brown, Sherri M.  
; APPLICANT: Dean, Duff A.  
; APPLICANT: Fromm, Michael E.  
; TITLE OF INVENTION: Synthetic DNA Sequences Having Enhanced  
; TITLE OF INVENTION: Expression in Monocytoidonous Plants and Method For  
; TITLE OF INVENTION: Preparation Thereof  
; NUMBER OF SEQUENCES: 164  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BB4F  
; STREET: 700 Chesterfield Parkway No. 5689052th  
; CITY: St. Louis  
; STATE: Missouri  
; COUNTRY: USA  
; ZIP: 63198

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,492  
FILING DATE:  
CLASSIFICATION: 800  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/172,333  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10605)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1919 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (synthetic)  
US-08-530-492-106

Query Match 78.2%; Score 17.2; DB 1; Length 1919;  
Best Local Similarity 86.4%; Pred. No. 20;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GGTCATCATCCAGATGAGAG 22  
Db 1315 GGTCATCATCCGATGAGAG 1294

## RESULT 2

US-08-906-517-106/c  
; Sequence 106, Application US/08906517  
; Patent No. 6180774  
; GENERAL INFORMATION:  
; APPLICANT: Brown, Sherri M.  
; APPLICANT: Dean, Duff A.  
; APPLICANT: Fromm, Michael E.  
; APPLICANT: Sanders, Patricia R.  
; TITLE OF INVENTION: Synthetic DNA Sequences Having Enhanced  
; TITLE OF INVENTION: Expression in Monocotyledonous Plants and Method For  
; TITLE OF INVENTION: Preparation Thereof  
; NUMBER OF SEQUENCES: 164  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/906,517  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: MOBT:170  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 512-418-3000  
; TELEFAX: 512-474-7577  
; INFORMATION FOR SEQ ID NO: 106:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1919 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-906-517-106

Query Match 78.2%; Score 17.2; DB 4; Length 1919;  
Best Local Similarity 86.4%; Pred. No. 20;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATCCAGATGAGAG 22  
Db 1315 GGTCATCATCCGATGAGAG 1294

## RESULT 3

US-09-052-089A-8/c  
; Sequence 8, Application US/09052089A  
; Patent No. 6345605  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Soo Y.  
; APPLICANT: Choi, Yongsun  
; TITLE OF INVENTION: SIGNAL TRANSDUCER FOR THE TNF RECEPTOR SUPER  
; FAMILY, AND USES THEREOF  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David A. Jackson, Esq.  
; STREET: 411 Hackensack Ave, Continental Plaza, 4th

Floor  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/052,089A  
; FILING DATE: 31-Mar-1998  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-198 CIP 1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-487-5800  
; TELEFAX: 201-343-1684  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1975 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: mouse  
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
US-09-052-089A-8

Query Match 78.2%; Score 17.2; DB 4; Length 1975;  
Best Local Similarity 86.4%; Pred. No. 20;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATCCAGATGAGAG 22  
Db 1294 GGACAGCATCCGATGAGAG 1273

## RESULT 4

US-09-770-315-4  
; Sequence 4, Application US/09770315  
; Patent No. 6429001  
; GENERAL INFORMATION:  
; APPLICANT: Chiron Corporation  
; TITLE OF INVENTION: Recombinant AAV Packaging Systems  
; FILE REFERENCE: 20263-501  
; CURRENT APPLICATION NUMBER: US/09/770,315  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: US 60/178,536  
; PRIOR FILING DATE: 2000-01-26  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 4072  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: recombinant DNA  
US-09-770-315-4

Query Match 78.2%; Score 17.2; DB 4; Length 4072;  
Best Local Similarity 86.4%; Pred. No. 23;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATCCAGATGAGAG 22  
Db 3151 GGTCATCATACAGACGAGAG 3172



; PRIOR APPLICATION NUMBER: US 60/177  
 ; PRIOR FILING DATE: 2000-01-26

```

; OPERATING SYSTEM: PC DOS/MS D
; SOFTWARE: PatentIn Release #1
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/652,971  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 225-3216  
TELEFAX: (415) 952-9881  
TELEX: 910 371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-652-971-1

Query Match 77.3%; Score 17; DB 1; Length 5769;  
Best Local Similarity 89.5%; Pred. No. 30;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 CATCATCCAGATGAAGAG 22  
||||| |||||||  
DB 4735 CATCATCCAGATGARGAG 4753

RESULT 12  
US-08-991-258A-1  
Sequence 1, Application US/08991258A  
Patent No. 5928887  
GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT, LLP  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/991,258A  
FILING DATE: 17-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/652,971  
FILING DATE: 24-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-63478-3/WHD/MTK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-991-258A-1

Query Match 77.3%; Score 17; DB 2; Length 5769;  
Best Local Similarity 89.5%; Pred. No. 30;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 CATCATCCAGATGAAGAG 22  
||||| |||||||  
DB 4735 CATCATCCAGATGARGAG 4753

RESULT 13  
US-08-769-399-1  
Sequence 1, Application US/08769399  
Patent No. 5976852  
GENERAL INFORMATION:  
APPLICANT: Cheng, Jill  
APPLICANT: Lasky, Laurence A.  
TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd.  
CITY: South San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/769,399  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: P1033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 225-3216  
TELEFAX: (415) 952-9881  
TELEX: 910 371-7168  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 379..4686  
US-08-769-399-1

Query Match 77.3%; Score 17; DB 2; Length 5769;  
Best Local Similarity 89.5%; Pred. No. 30;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 CATCATCCAGATGAAGAG 22  
||||| |||||||  
DB 4735 CATCATCCAGATGARGAG 4753

RESULT 14

US-08-991-953A-1  
; Sequence 1, Application US/08991953A  
; Patent No. 6083748  
; GENERAL INFORMATION:  
; APPLICANT: Cheng, Jill  
; APPLICANT: Lasky, Laurence A.  
; TITLE OF INVENTION: A NOVEL KAPPA/MU-LIKE PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE, PTP LAMBDA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT, LLP  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING-SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; APPLICATION NUMBER: US/08/991,953A  
; FILING DATE: 16-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/652,971  
; FILING DATE: 24-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Walter H.  
; REGISTRATION NUMBER: 24,190  
; REFERENCE/DOCKET NUMBER: A-63478-3/WHID/MTK  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5769 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 379..4686  
US-08-991-953A-1

Query Match 77.3%; Score 17; DB 3; Length 5769;  
Best Local Similarity 89.5%; Pred. No. 30;  
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 CATCATTCAGATGAAG 22  
DB 4735 CATCATTCAGATGAAG 4753  
||||| ||||| ||||| |||||

RESULT 15  
US-08-611-107-30/c  
; Sequence 30, Application US/08611107  
; Patent No. 5801233  
; GENERAL INFORMATION:  
; APPLICANT: Haselkorn, Robert  
; APPLICANT: Gornicki, Piotr  
; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING  
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE AND USES  
; TITLE OF INVENTION: THEREFOR  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas

; COUNTRY: United States of America  
; ZIP: 77210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING-SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/611,107  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US SN 07/956,700  
; FILING DATE: 02-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US PCT/US93/09340  
; FILING DATE: 30-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US SN 08/422,560  
; FILING DATE: 14-APR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kitchell, Barbara S.  
; REGISTRATION NUMBER: 33,928  
; REFERENCE/DOCKET NUMBER: ARCD:221  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (512) 474-7577  
; INFORMATION FOR SEQ ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 11748 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-611-107-30  
Query Match 73.6%; Score 16.2; DB 1; Length 11748;  
Best Local Similarity 85.7%; Pred. No. 81;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GGTCATCATTCAGATGAAGA 21  
DB 10950 GATCATCTTCCAGGTGAAGA 10930  
| ||||| ||||| ||||| |||||  
Search completed: July 8, 2003, 09:31:47  
Job time : 31.4207 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 138.224 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-12

Perfect score: 22

Sequence: 1 ggtcatcattccagatgaagag 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	AAZ28818	Rat membrane metal
c 2	22	100.0	2765	AAZ28810	Rat membrane metal
c 3	20.4	92.7	2286	24 AAD28130	Soluble secreted e
c 4	20.4	92.7	2925	21 AAF63763	CDNA encoding neut
c 5	18.8	85.5	2076	22 AAF89737	Nucleotide sequenc
c 6	18.8	85.5	2332	24 AAS97186	Human metalloprote
c 7	18.8	85.5	2262	22 AAF89739	Nucleotide sequenc
c 8	18.8	85.5	2318	24 AAD30580	Human protease, PR
c 9	18.8	85.5	2340	22 AAF89738	Nucleotide sequenc

c 10	18.8	85.5	2580	24 ABN84280	Human SEP endopept
c 11	18.8	85.5	2636	22 AAF59660	Human neprilysin-1
c 12	18.8	85.5	2663	22 AAF59661	Human neprilysin-1
c 13	18.8	85.5	2676	21 AAF63764	CDNA encoding neut
c 14	18.8	85.5	2714	22 AAF59659	Human neprilysin-1
c 15	18.8	85.5	2893	24 ABN84279	Human SEP endopept
c 16	18.8	85.5	2893	24 AAD28544	Human soluble secr
c 17	18.8	85.5	2953	24 ABK48251	CDNA encoding nove
c 18	18.8	85.5	2975	24 AAD28547	Human SEP CDNA inc
c 19	18.8	81.8	645	24 ABK33029	DNA encoding novel
c 20	17.2	78.2	528	22 AAH66892	C glutamicum codin
c 21	17.2	78.2	1260	20 AAZ35940	Streptococcus pneu
c 22	17.2	78.2	1260	20 AAZ35941	Streptococcus pneu
c 23	17.2	78.2	1905	11 AAQ05798	Synthetic B.t.k. p
c 24	17.2	78.2	1905	22 AA508952	Bacillus thuringie
c 25	17.2	78.2	1907	21 AAAL5561	Cry2Aa delta-endot
c 26	17.2	78.2	1919	19 AAV00396	Insecticidal gene
c 27	17.2	78.2	1919	22 AAF73312	CryIIA insecticida
c 28	17.2	78.2	1969	22 ABK45341	Human breast cell
c 29	17.2	78.2	1969	22 ABK55830	Human foetal liver
c 30	17.2	78.2	1969	22 ABA25506	Probe #3972 for ge
c 31	17.2	78.2	1969	22 AKO4048	Human brain expres
c 32	17.2	78.2	1969	22 AAQ29533	Human bone marrow
c 33	17.2	78.2	1969	22 AAI14105	Probe #4038 for ge
c 34	17.2	78.2	1969	22 AAI35486	Probe #4172 used t
c 35	17.2	78.2	1969	22 AAI03958	Probe #3949 used t
c 36	17.2	78.2	1969	24 ABS04084	Human genome-deriv
c 37	17.2	78.2	1975	24 AAD31619	Mouse TNF receptor
c 38	17.2	78.2	4072	22 AAH26327	Adenovirus helper
c 39	17.2	78.2	4675	22 AAH41481	Adeno-associated v
c 40	17.2	78.2	4675	22 AAF89931	Nucleotide sequenc
c 41	17.2	78.2	4675	22 ABA02989	Adeno-associated v
c 42	17.2	78.2	4679	22 AAI66974	Adeno-associated v
c 43	17.2	78.2	4679	22 AAF23750	AAV2 DNA sequence.
c 44	17.2	78.2	4680	17 AAT09008	Wild-type adeno-as
c 45	17.2	78.2	7557	22 AAH26326	Adeno-associated v

#### ALIGNMENTS

RESULT 1  
AAZ28818  
ID AAZ28818 standard; DNA; 22 BP.  
XX  
AC AAZ28818;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Rat membrane metalloprotease NEPII gene probe #8.  
XX  
KW Rat; membrane metalloprotease; neprilysin II; NEPII; inactivation; ss;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
XX  
OS Synthetic.  
OS Rattus rattus.  
XX  
PN FR2777291-AL.  
XX  
PD 15-OCT-1999.  
XX  
PF 08-APR-1998; 98FR-0004389.  
XX  
PR 08-APR-1998; 98FR-0004389.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
XX  
DR WPI; 1999-593429/51.

XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease 3  
XX  
XX Claim 3; Page 22; 29pp; French.  
XX Sequences AA228811-228827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysine II (NEPII) gene.  
CC (AA228810). NEPII is involved in (in)activation of neuronal and hormonal  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
CC disorders.  
XX  
XX Sequence 22 BP; 7 A; 4 C; 6 G; 5 T; 0 other;  
XX  
XX Query Match 100.0%; Score 22; DB 20; Length 22;  
XX Best Local Similarity 100.0%; Pred. No. 0.33; 0; Indels 0; Gaps 0;  
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 GGTCATCATTCAGATGAAGAG 22  
XX |||||  
XX Db 1 GGTCATCATTCAGATGAAGAG 22  
XX |||||  
XX  
XX RESULT 2  
XX AA228810/c  
XX ID AA228810 standard; cDNA; 2765 BP.  
XX  
XX AC AA228810;  
XX  
XX DT 01-FEB-2000 (first entry)  
XX  
XX DE Rat membrane metalloprotease NEPII gene.  
XX  
XX KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
XX neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
XX cardiovascular disease; neurodegenerative disease; growth disorder;  
XX hypothalamic-hypophyseal axis; endocrine disorder; ds.  
XX  
XX OS Rattus rattus.  
XX  
XX PN FR2777291-A1.  
XX  
XX PD 15-OCT-1999.  
XX  
XX PF 08-APR-1998; 98FR-0004389.  
XX  
XX PR 08-APR-1998; 98FR-0004389.  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX PA Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
XX Schwartz JC;  
XX  
XX DR WPI; 1999-593429/51.  
XX P-PSDB; AA44177.  
XX  
XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease -  
XX  
XX Claim 2; Page 12-16; 29pp; French.  
XX  
XX This sequence represents the gene for the rat membrane metalloprotease  
CC designated neprilysine II (NEPII), which is involved in (in)activation  
CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
CC for specific substrates (used to detect NEPII in cells and tissues) or  
CC inhibitors, which can also be used to detect NEPII or for treatment of

CC disorders related to peptidergic signalling in which NEPII is involved,  
CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
CC endocrine disorders.  
XX  
XX Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
XX  
XX Query Match 100.0%; Score 22; DB 20; Length 2765;  
XX Best Local Similarity 100.0%; Pred. No. 0.69; 0; Indels 0; Gaps 0;  
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 GGTCATCATTCAGATGAAGAG 22  
XX |||||  
XX Db 818 GGTCATCATTCAGATGAAGAG 797  
XX |||||  
XX  
XX RESULT 3  
XX AAD28130/c  
XX ID AAD28130 standard; DNA; 2286 BP.  
XX  
XX AC AAD28130;  
XX  
XX DT 07-MAY-2002 (first entry)  
XX  
XX DE Soluble secreted endopeptidase (SEP) consensus DNA.  
XX  
XX KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
XX enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
XX MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
XX FSD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
XX  
XX OS Homo sapiens.  
XX Mus sp.  
XX OS Rattus sp.  
XX  
XX FH Key Location/Qualifiers  
XX FT misc\_feature 1664..2286  
XX FT /\*tag= "a  
XX FT /note= "Encodes catalytic domain"  
XX  
XX PN WO200206492-A1.  
XX  
XX PD 24-JAN-2002.  
XX  
XX PF 16-JUL-2001; 2001WO-IB01263.  
XX  
XX PR 14-JUL-2000; 2000GB-0017387.  
XX  
XX PA (PFIZ ) PFIZER LTD.  
XX PA (PFIZ ) PFIZER INC.  
XX  
XX PI Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
XX  
XX DR WPI; 2002-155042/20.  
XX  
XX An isolated and/or purified nucleic acid encoding a human soluble  
PT secreted endopeptidase which is useful for treating sexual dysfunction,  
PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
PT female sexual arousal disorder -  
XX  
XX PS Disclosure; Fig 6; 167pp; English.  
XX  
XX The invention relates to an isolated and/or purified nucleic acid  
CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
CC antibody and the compound which inhibits or selectively inhibits the  
CC human SEP protein are useful in the manufacture of a medicament for  
CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
CC also useful for treating the above disorders and other disorders such  
CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
CC hyposexual sexual desire disorder. The present sequence is SEP consensus  
CC DNA sequence found in human, mouse and rat.

XX SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;

Query Match 92.7%; Score 20.4; DB 24; Length 2286;  
Best Local Similarity 95.5%; Pred. No. 4;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGTCAATCATTCAGATGAAGAG 22  
||||| ||||||| ||||||| |||||||  
DB 694 GGTCTCATTCAGATGAAGAG 673

RESULT 4  
ID AAA63763/C  
XX ID AAA63763 standard; CDNA; 2925 BP.  
AC AAA63763;  
XX  
DT 04-DEC-2000 (first entry)  
XX  
DE CDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.  
XX  
KW Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
KW NEP-like enzyme; protein production; protein secretion;  
KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
KW fertility; bone disease; abnormal phosphate metabolism; ss.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 332..2629  
FT FT /\*tag= a  
FT FT /product= "neutral endopeptidase metalloproteinase-like  
FT FT enzyme NL-1"  
XX  
PN WO200047750-A2.  
XX  
PD 17-AUG-2000.  
XX  
PF 11-FEB-2000; 2000WO-CA00147.  
XX  
PR 11-FEB-1999; 99CA-2260376.  
XX  
PA (UIMO-) UNIV MONTREAL.  
XX  
PI Desgroseillers L, Boileau G;  
XX  
XX WPI; 2000-549148/50.  
DR P-PSDB; AAB08130.  
XX  
XX Novel neutral endopeptidase-like metalloproteinase polypeptides and  
FT polynucleotides, used to screen for related sequences and enzyme  
FT inhibitors, used for the treatment of NL-3 related bone disorders -  
XX  
XX Disclosure; Fig 3; 59pp; English.  
PS  
XX The present sequence encodes a murine neutral endopeptidase  
XX metalloproteinase-like enzyme, designated NL-1. The specification  
CC also describes NL-2 and NL-3. The NL enzymes are used to test for  
CC specific inhibitors. The N-terminal region of the enzymes can be used  
CC to promote production and secretion of foreign proteins and active  
CC biopeptides, using chimeric constructs containing the foreign protein  
CC downstream from and in phase with the N-terminal region. The NL enzymes  
CC have been localised to the brain, and may be useful in the  
CC treatment of neurological diseases such as Alzheimer's disease, pain,  
CC and psychiatric disorders. NL enzymes have also been localised to the  
CC testis and ovaries, and may be used to control fertility. They have  
CC also been localised to bones, and may be used to treat bone diseases,  
CC and abnormal phosphate metabolisms related to improper peptide  
XX processing by the NL-3 enzyme.  
XX  
SQ Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;

Query Match 92.7%; Score 20.4; DB 21; Length 2925;  
Best Local Similarity 95.5%; Pred. No. 4.2;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGTCAATCATTCAGATGAAGAG 22  
||||| ||||||| ||||||| |||||||  
DB 1013 GGTCTCATTCAGATGAAGAG 992

RESULT 5  
ID AAF89737/C  
XX ID AAF89737 standard; DNA; 2076 BP.  
AC AAF89737;  
XX  
DT 23-JUL-2001 (first entry)  
XX  
DE Nucleotide sequence of a human metalloproteinase enzyme IGS5.  
XX  
KW Metalloproteinase; IGS5; infection; pain; cancer; diabetes; obesity;  
KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
KW hypotension; hypertension; urinary retention; osteoporosis;  
KW angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
KW benign prostatic hypertrophy; migraine; psychotic disorder;  
KW neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
KW neurodegenerative disease; sleep disorder; epilepsy; kidney disease;  
KW cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
KW subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
KW peripheral vascular disease; Raynaud's disease; motility disorder;  
KW gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
KW inflammation; chemotherapy induced injury; tumour invasion;  
KW immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
KW severe mental retardation; dyskinesia; Huntington's disease;  
KW Gilles de la Tourette's syndrome; ss.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2076  
FT FT /\*tag= a  
FT FT /product= "metalloproteinase enzyme IGS5"  
XX  
XX WO200136610-A1.  
XX  
PD 25-MAY-2001.  
XX  
PF 17-NOV-2000; 2000WO-EP11532.  
XX  
PR 19-NOV-1999; 99EP-0203862.  
PR 19-NOV-1999; 99NL-1013616.  
PR 31-MAY-2000; 2000EP-0201937.  
PR 31-MAY-2000; 2000NL-1015356.  
XX  
PA (SOLV ) SOLVAY PHARM BV.  
XX  
XX Deleersnijder W, Wieggers R, Weske M;  
PI WPI; 2001-343815/36.  
XX  
DR P-PSDB; AAB83840.  
XX  
XX New IGS5 polypeptides useful for treating infections, pain, cancer,  
PT diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
PT hypertension, urinary retention and Parkinson's disease  
XX  
XX Claim 11; Page 5-6; 115pp; English.  
PS  
XX The present sequence encodes a human metalloproteinase enzyme designated  
CC IGS5. IGS5 polynucleotides and polypeptides are useful for treating  
CC infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
CC urinary retention, osteoporosis, angina pectoris, myocardial infarction,  
CC urinary ulcers, allergies, benign prostatic hypertrophy, migraine,  
CC psychotic and neurological disorders, autism, multiple sclerosis,

CC Alzheimer's disease, and other neurodegenerative diseases, sleep  
 CC disorder, epilepsy, cardiovascular diseases, arteriosclerosis,  
 CC cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral  
 CC infarction, peripheral vascular disease, Raynaud's disease, kidney  
 CC diseases, gastrointestinal disorders, motility disorders and conditions  
 CC of delayed gastric emptying, post-operative or diabetic gastroparesis,  
 CC diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
 CC immune disorders, arthritis, endotoxin shock, sepsis, complications of  
 CC diabetes mellitus, and severe mental retardation and dyskinesias, such  
 CC as Huntington's disease or Gilles de la Tourette's syndrome.

XX  
 SQ Sequence 2076 BP; 493 A; 578 C; 636 G; 369 T; 0 other;

Query Match 85.5%; Score 18.8; DB 22; Length 2076;  
 Best Local Similarity 90.9%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| || ||||| ||||| |||||  
 Db 460 GGTCTGCTGTCAGATGAAGAG 439

## RESULT 6

AAS97186/c  
 ID AAS97186 standard; cDNA; 2232 BP.

XX  
 AC AAS97186;

XX  
 DT 26-FEB-2002 (first entry)

XX  
 DE Human metalloprotease partial DNA sequence #15.

XX Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;  
 KW vasotropic; antimigraine; analgesic; endocrine; neurotropic; tranquiliser;  
 KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;  
 KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;  
 KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;  
 KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;  
 KW immune-related disease; cardiovascular disease; neuronal disease;  
 KW migraine; sexual dysfunction; mood disorder; attention disorder;  
 KW cognition disorder; hypotension; hypertension; psychotic disorder;  
 KW dyskinesia; metabolic disorder; inflammatory disorder; ss.

XX Homo sapiens.

XX WO200183782-A2.

XX 08-NOV-2001.

XX 04-MAY-2001; 2001WO-US14431.

XX 04-MAY-2000; 2000US-201879P.

XX (SUGGE-) SUGEN INC.

XX Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
 PI Payne V;

XX WPI; 2002-041502/05.

XX P-PSDB; AAU72903.

XX Novel protease polypeptide useful for screening for substances that may  
 PT be used to treat, e.g., cancers, immune-related diseases,  
 PT cardiovascular disease, migraine, pain, psychotic and inflammatory  
 PT disorders -

XX Claim 30; Figure 1R-S; 232pp; English.

XX The invention relates to an isolated, enriched, or purified protease  
 CC polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to  
 CC screen for substances (S) that may modulate its activity. Administering  
 CC S (which modulates protease activity in vitro) may be used to treat a  
 CC disease or disorder selected from cancers (e.g., of tissues, of blood or

CC haematopoietic origin, of the breast, colon, lung, prostate, cervical,  
 CC brain, ovarian, bladder or kidney), immune-related diseases and  
 CC cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral  
 CC infarction, peripheral vascular disease, Raynaud's disease, kidney  
 CC diseases, gastrointestinal disorders, motility disorders and conditions  
 CC of delayed gastric emptying, post-operative or diabetic gastroparesis,  
 CC diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
 CC immune disorders, arthritis, endotoxin shock, sepsis, complications of  
 CC diabetes mellitus, and severe mental retardation and dyskinesias, such  
 CC as Huntington's disease or Gilles de la Tourette's syndrome.

XX  
 SQ Sequence 2232 BP; 512 A; 620 C; 705 G; 395 T; 0 other;

Query Match 85.5%; Score 18.8; DB 24; Length 2232;  
 Best Local Similarity 90.9%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| || ||||| ||||| |||||  
 Db 619 GGTCTGCTGTCAGATGAAGAG 598

## RESULT 7

AAF89739/c

ID AAF89739 standard; DNA; 2262 BP.

XX  
 AC AAF89739;

XX  
 DT 23-JUL-2001 (first entry)

XX  
 DE Nucleotide sequence of a human metalloprotease enzyme IGS5.

XX Metalloprotease; IGS5; infection; pain; cancer; diabetes; obesity;  
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
 KW hypotension; hypertension; urinary retention; osteoporosis;  
 KW angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
 KW benign prostatic hypertrophy; migraine; psychotic disorder;  
 KW neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
 KW neurodegenerative disease; sleep disorder; epilepsy; kidney disease;  
 KW cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
 KW subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
 KW peripheral vascular disease; Raynaud's disease; motility disorder;  
 KW gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
 KW inflammation; chemotherapy induced injury; tumour invasion;  
 KW immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
 KW severe mental retardation; dyskinesia; Huntington's disease;  
 KW Gilles de la Tourette's syndrome; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..2262

FT /\*tag= a

FT /product= "metalloprotease enzyme IGS5"

XX WO200136610-A1.

XX 25-MAY-2001.

XX 17-NOV-2000; 2000WO-EPI1532.

XX 19-NOV-1999; 99EP-0203862.

XX 19-NOV-1999; 99NL-1013616.

XX 31-MAY-2000; 2000EP-0201937.

XX 31-MAY-2000; 2000NL-1015356.

XX (SOLV.) SOLVAY PHARM BV.

XX Deleersnijder W, Wiegers R, Weske M;

XX WPI; 2001-343815/36.

XX P-PSDB; AAB83842.

PT New IGS5 polypeptides useful for treating infections, pain, cancer,  
 PT diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
 PT hypertension, urinary retention and Parkinson's disease  
 XX  
 PS Claim 11; Page 8-9; 115pp; English.  
 XX

CC The present sequence encodes a human metalloprotease enzyme designated  
 CC IGS5. IGS5 polynucleotides and polypeptides are useful for treating  
 CC infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, osteoporosis, angina pectoris, myocardial infarction,  
 CC stroke, ulcers, allergies, benign prostatic hypertrophy, migraine,  
 CC psychotic and neurological disorders, autism, multiple sclerosis,  
 CC Alzheimer's disease, and other neurodegenerative diseases, sleep  
 CC disorder, epilepsy, cardiovascular diseases, arteriosclerosis, cerebral  
 CC cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral  
 CC infarction, peripheral vascular disease, Raynaud's disease, kidney  
 CC diseases, gastrointestinal disorders, motility disorders and conditions  
 CC of delayed gastric emptying, post-operative or diabetic gastroparesis,  
 CC diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
 CC immune disorders, arthritis, endotoxin shock, sepsis, complications of  
 CC diabetes mellitus, and severe mental retardation and dyskinesias, such  
 CC as Huntington's disease or Gilles de la Tourette's syndrome.  
 XX

XX Sequence 2262 BP; 520 A; 628 C; 716 G; 398 T; 0 other;

Query Match 85.5%; Score 18.8; DB 22; Length 2262;

Best Local Similarity 90.9%; Pred. No. 24;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGTCATCATTCACAGATGAAGAG 22

DB 646 GGTCTGCTTCACAGATGAAGAG 625

RESULT 8

AAD30580/C

ID AAD30580 standard; cDNA; 2318 BP.

XX AAD30580;

XX 21-MAY-2002 (first entry)

XX Human protease, PRTS-13 cDNA.

XX Human; protease; PRTS-13; enzyme; gastritis; cirrhosis; Crohn's disease;  
 KW gastrointestinal disorder; autoimmune; inflammatory; cell proliferative;  
 KW cardiovascular; developmental; epithelial; neurological; reproductive;  
 KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;  
 KW anaemia; asthma; atherosclerosis; hypertension; myocardial infarction;  
 KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;  
 KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;  
 KW Pick's disease; infertility; vitiligo; drug screening; gene therapy; ss.  
 XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 338..1651

FT /\*tag= a

FT /product= "Human PRTS-13 protein"

FT sig\_peptide 338..427

FT /\*tag= b

FT mat\_peptide 428..1648

FT /\*tag= c

FT /product= "Mature PRTS-13 protein"

FT WO200208396-A2.

XX 31-JAN-2002.

XX 17-JUL-2001; 2001WO-US22397.

XX 21-JUL-2000; 2000US-220063P.

PR 28-JUL-2000; 2000US-221680P.  
 PR 04-AUG-2000; 2000US-223544P.  
 PR 11-AUG-2000; 2000US-224717P.  
 PR 16-AUG-2000; 2000US-225988P.  
 PR 23-AUG-2000; 2000US-227568P.

XX (INCY-) INCYTE GENOMICS INC.

XX Deleage AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;

PI Tribouley CM; Das D, Kallick DA, Nguyen DB, Lee EA, Khan FA;

PI Yue H, Au-Young J, Griffin JA, Pollick JL, Ramkumar J, Yang J;

PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;

PI Sanjanwala MS, Yao MG, Burford N, Walla NK, Lal P, Lee S, Todd S;

PI Lo TP, Tang YT, Elliott VS, Azimzal Y, Lu Y;

XX WPI; 2002-206082/26.

DR P-PSDB; AAE19176.

XX New human protease polypeptide, useful in diagnosis, prevention and

DR treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,

XX cell proliferative, developmental, epithelial and neurological

PT disorders

PT Claim 5; Page 174-175; 182pp; English.

XX The invention relates to an isolated human protease polypeptide (PRTS).

XX PRTS protein and DNA are useful for diagnosing, treating and preventing

CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),

CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,

CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,

CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,

CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),

CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders

CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,

CC Parkinson's disease), and reproductive disorders (infertility). PRTS

CC protein is useful in a number of drug screening techniques and to

CC analyse the proteome of a tissue or cell type. PRTS DNA is useful for

CC creating knockin humanised animals or transgenic animals to model human

CC diseases, in somatic or germline gene therapy and in microarrays

CC utilising fluids or tissues from patients to detect altered PKIN

CC expression. The present sequence is human PRTS-13 cDNA.

XX Sequence 2318 BP; 526 A; 647 C; 735 G; 410 T; 0 other;

SQ Query Match 85.5%; Score 18.8; DB 24; Length 2318;

Best Local Similarity 90.9%; Pred. No. 24;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGTCATCATTCACAGATGAAGAG 22

DB 708 GGTCTGCTTCACAGATGAAGAG 687

RESULT 9

AAF89738/C

ID AAF89738 standard; DNA; 2340 BP.

XX AAF89738;

XX 23-JUL-2001 (first entry)

XX Nucleotide sequence of a human metalloprotease enzyme IGS5.

XX Metalloprotease; IGS5; infection; pain; cancer; diabetes; obesity;

XX anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;

XX hypotension; hypertension; urinary retention; osteoporosis;

XX angina pectoris; myocardial infarction; stroke; ulcer; allergy;

XX benign prostatic hypertrophy; migraine; psychotic disorder;

XX neurological disorder; autism; multiple sclerosis; Alzheimer's disease;

XX neurodegenerative disease; sleep disorder; epilepsy; kidney disease;

XX cardiovascular disease; arteriosclerosis; cerebrovasospasm;

XX subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;

XX peripheral vascular disease; Raynaud's disease; motility disorder;



Db 973 GGTCGTCGTTCCAGATGAAGAG 952

RESULT 11  
AAF59660/c  
ID AAF59660 standard; cDNA: 2636 BP.  
XX AAF59660;  
XX  
XX  
XX 27-APR-2001 (first entry)  
XX  
XX Human neprilysin-like membrane metallopeptidase SNEPb cDNA.  
XX  
XX Human; SNEPb; neprilysin-like membrane metallopeptidase;  
KW splice variant; alternative splicing; zinc endopeptidase family;  
KW neuropeptide; peptide hormone; processing; metabolism; vaccine;  
KW drug screening; renal insufficiency; hepatic disease; ischaemia; pain;  
KW hypertension; cancer; inflammation; cardiovascular disease;  
KW neuronal disease; pancreatic disease; prostatic disease;  
KW respiratory disease; gene therapy; nephrotropic; vasotropic; analgesic;  
KW cerebroprotective; hypotensive; cytostatic; antiinflammatory;  
KW cardiovascular; hepatotropic; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1069188-A1.  
XX  
XX 17-JAN-2001.  
XX  
XX 15-JUL-1999; 99EP-0401767.  
XX  
XX 15-JUL-1999; 99EP-0401767.  
XX (SNFI ) SANOFI-SYNTHELABO.  
XX  
XX Jagerschmidt A, Agnel M, Culouscou J;  
XX  
XX WPI: 2001-212582/22.  
XX P-PSDB; AAB60562.  
XX  
XX New membrane-associated metallopeptidase SNEPA, SNEPb and SNEPC  
PT polypeptides and polynucleotides, useful for treating e.g. acute and  
PT chronic renal insufficiency, pain, stroke, cancer, inflammation, renal  
PT and hepatic ischemia  
XX  
XX Claim 5; Page 30-33; 72pp; English.  
XX  
XX The invention relates to the human SNEPA, SNEPb and SNEPC proteins, and  
CC the cDNAs encoding them. SNEPA, SNEPb and SNEPC are neprilysin-like  
CC membrane metallopeptidases and are the products of alternative splicing.  
CC The substrate(s) for the SNEP proteins are not as yet known, although  
CC the neprilysin family of zinc endopeptidases play key roles in the  
CC processing and/or metabolism of neuropeptides and peptide hormones. SNEP  
CC nucleotides may be used as hybridisation probes for cDNA and genomic  
CC DNA; to isolate full-length cDNAs and genomic clones encoding SNEPA,  
CC SNEPb or SNEPC; to isolate cDNA and genomic clones of SNEP homologues;  
CC as research reagents and material for the discovery of treatments and  
CC diagnostics for animal and human diseases; and for chromosome  
CC identification. The SNEP proteins may be used as immunogens to  
CC produce antibodies immunospecific for SNEPA, SNEPb or SNEPC. Such  
CC antibodies are used to isolate or identify clones expressing the  
CC protein, or to purify the proteins by affinity chromatography.  
CC  
CC SNEP nucleotides, and SNEP activators or inhibitors may be used  
CC to treat acute and chronic renal insufficiency, cancer, inflammation, as  
CC ischaemia, pain, stroke, hypertensive disease, renal, hepatic  
CC as well as cardiovascular, neuronal, pancreatic, prostatic, renal,  
CC respiratory or hepatic diseases. they may also be used in modulating  
CC peptide activation and/or degradation in the brain or kidney or in  
CC another organ, or to diagnose or treat any disorder related to abnormal

CC expression of SNEPA, SNEPb or SNEPC. The present sequence represents  
CC cDNA encoding SNEPb.  
XX  
SQ Sequence 2636 BP; 601 A; 759 C; 813 G; 462 T; 1 other;  
Query Match 85.5%; Score 18.8; DB 22; Length 2636;  
Best Local Similarity 90.9%; Pred. No. 25;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GGTCATCATTCAGATGAAGAG 22  
||||| ||| ||||| ||||| |||||  
DB 653 GGTCGTCGTTCCAGATGAAGAG 632

RESULT 12  
AAF59661/c  
ID AAF59661 standard; cDNA: 2663 BP.  
XX AAF59661;  
XX  
XX 27-APR-2001 (first entry)  
XX  
XX Human neprilysin-like membrane metallopeptidase SNEPc cDNA.  
XX  
XX Human; SNEPc; neprilysin-like membrane metallopeptidase;  
KW splice variant; alternative splicing; zinc endopeptidase family;  
KW neuropeptide; peptide hormone; processing; metabolism; vaccine;  
KW drug screening; renal insufficiency; hepatic disease; ischaemia; pain;  
KW hypertension; cancer; inflammation; cardiovascular disease;  
KW neuronal disease; pancreatic disease; prostatic disease;  
KW respiratory disease; gene therapy; nephrotropic; vasotropic; analgesic;  
KW cerebroprotective; hypotensive; cytostatic; antiinflammatory;  
KW cardiovascular; hepatotropic; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1069188-A1.  
XX  
XX 17-JAN-2001.  
XX  
XX 15-JUL-1999; 99EP-0401767.  
XX  
XX 15-JUL-1999; 99EP-0401767.  
XX (SNFI ) SANOFI-SYNTHELABO.  
XX  
XX Jagerschmidt A, Agnel M, Culouscou J;  
XX  
XX WPI: 2001-212582/22.  
XX P-PSDB; AAB60563.  
XX  
XX New membrane-associated metallopeptidase SNEPA, SNEPb and SNEPC  
PT polypeptides and polynucleotides, useful for treating e.g. acute and  
PT chronic renal insufficiency, pain, stroke, cancer, inflammation, renal  
PT and hepatic ischemia  
XX  
XX Claim 5; Page 35-38; 72pp; English.  
XX  
XX The invention relates to the human SNEPA, SNEPb and SNEPC proteins, and  
CC the cDNAs encoding them. SNEPA, SNEPb and SNEPC are neprilysin-like  
CC membrane metallopeptidases and are the products of alternative splicing.  
CC The substrate(s) for the SNEP proteins are not as yet known, although  
CC the neprilysin family of zinc endopeptidases play key roles in the  
CC processing and/or metabolism of neuropeptides and peptide hormones. SNEP  
CC nucleotides may be used as hybridisation probes for cDNA and genomic  
CC DNA; to isolate full-length cDNAs and genomic clones encoding SNEPA,  
CC SNEPb or SNEPC; to isolate cDNA and genomic clones of SNEP homologues;  
CC as research reagents and material for the discovery of treatments and  
CC diagnostics for animal and human diseases; and for chromosome  
CC identification. The SNEP proteins may be used as immunogens to  
CC produce antibodies immunospecific for SNEPA, SNEPb or SNEPC. Such  
CC antibodies are used to isolate or identify clones expressing the  
CC protein, or to purify the proteins by affinity chromatography.



CC DNA; to isolate full-length cDNAs and genomic clones encoding SNEPa,  
 CC SNEPb or SNEPc; to isolate cDNA and genomic clones of SNEP homologues;  
 CC as research reagents and material for the discovery of treatments and  
 CC diagnostics for animal and human diseases; and for chromosome  
 CC identification. The SNEP proteins may be used as immunogens to  
 CC produce antibodies immunospecific for SNEPa, SNEPb or SNEPc. Such  
 CC antibodies are used to isolate or identify clones expressing the  
 CC protein, or to purify the proteins by affinity chromatography.  
 CC SNEP proteins may also be used in screening for compounds which modulate  
 CC SNEP endopeptidase activity, and to assess enzymatic cleavage of small  
 CC molecule substrates in cells, cell-free preparations, chemical libraries  
 CC and product mixtures. The SNEP proteins (as vaccine compositions),  
 CC SNEP nucleotides, and SNEP activators or inhibitors may used  
 CC to treat acute and chronic renal insufficiency, renal and hepatic  
 CC ischaemia, pain, stroke, hypertensive disease, cancer, inflammation, as  
 CC well as cardiovascular, neuronal, pancreatic, prostatic, renal,  
 CC respiratory or hepatic diseases. They may also be used in modulating  
 CC peptide activation and/or degradation in the brain or kidney or in  
 CC another organ, or to diagnose or treat any disorder related to abnormal  
 CC expression of SNEPa, SNEPb or SNEPc. The present sequence represents  
 CC cDNA encoding SNEPa.

XX SQ Sequence 2714 BP; 621 A; 780 C; 836 G; 476 T; 1 other;

Query Match 85.5%; Score 18.8; DB 22; Length 2714;  
 Best Local Similarity 90.9%; Pred. No. 25;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GGTCATCATTCACAGATGAAGAG 22  
 ||||| || ||||| ||||| |||||  
 Db 731 GGTCGTCGTCACAGATGAAGAG 710.

## RESULT 15

ABN84279/c  
 ID ABN84279 standard; cDNA; 2893 BP.

XX AC ABN84279;

XX DT 23-SEP-2002 (first entry)

XX DE Human SEP endopeptidase coding sequence.

XX KW SEP; endopeptidase; enzyme; human; neuropeptide Y1; receptor;  
 KW male sexual dysfunction; male erectile dysfunction; obesity;  
 KW anorexia; bulimia; vasotropic; anorectic; therapy; inhibitor; gene;  
 KW ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 FT CDS 185..2524  
 FT /\*tag= a  
 FT /product= "SEP"

XX PN WO200247670-A1.

XX PD 20-JUN-2002.

XX PF 10-DEC-2001; 2001WO-1B02399.

XX PR 15-DEC-2000; 2000GB-0030647.

XX PR 06-APR-2001; 2001GB-0008730.

XX PR 23-APR-2001; 2001GB-0009910.

XX PR 04-MAY-2001; 2001GB-0011037.

XX PR 29-JUN-2001; 2001US-0895367.

XX PR 13-JUL-2001; 2001US-0905846.

XX PR 24-AUG-2001; 2001GB-0020679.

XX (PF12 ) PFIZER LTD.

XX PA (PF12 ) PFIZER INC.

XX PI Naylor AM, Van Der Graaf PH, Wayman CP;

XX

DR WPI; 2002-547828/58.

DR P-PSDB; ABB79521.

XX

PT Use of an inhibitor of neuropeptide Y in the preparation of medicament  
 for the treatment or prevention of male erectile dysfunction

XX

PS Disclosure; Fig 7; 179pp; English.

XX

CC The present sequence is a nucleotide sequence coding for human  
 CC SEP, a soluble secreted endopeptidase. The invention relates to  
 CC the use of an inhibitor of NPY, or an inhibitor of NPY Y1 receptor,  
 CC especially an inhibitor selective for an NPY or NPY Y1 receptor  
 CC associated with male genitalia, in the preparation of a medicament  
 CC for the treatment or prevention of male sexual dysfunction,  
 CC especially male erectile dysfunction (MED). The NPY inhibitor  
 CC may be used with an auxiliary active agent such as an SEP  
 CC inhibitor. The invention provides a SEP assay that can be used to  
 CC detect candidate inhibitors of SEP. In addition to treatment of  
 CC MED, NPY inhibitors can also be used to treat abnormal drink and  
 CC food intake disorders, such as obesity, bulimia, anorexia and  
 CC metabolic disorders.

XX SQ Sequence 2893 BP; 658 A; 854 C; 876 G; 505 T; 0 other;

Query Match 85.5%; Score 18.8; DB 24; Length 2893;  
 Best Local Similarity 90.9%; Pred. No. 25;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGTCATCATTCACAGATGAAGAG 22

||||| || ||||| ||||| |||||

Db 908 GGTCGTCGTCACAGATGAAGAG 887

Search completed: July 8, 2003, 02:18:59

Job time : 139.224 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 127.063 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-12

Perfect score: 22

Sequence: 1 ggtcatcattccagatgaagag 22

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published Applications\_NA.\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.8	85.5	2893	9 US-10-017-273A-4	Sequence 4, Appli
C 2	18.8	85.5	2893	10 US-09-905-846-1	Sequence 1, Appli
C 3	18.8	85.5	2975	9 US-10-017-273A-5	Sequence 5, Appli
C 4	18.8	85.5	2975	10 US-09-905-846-5	Sequence 5, Appli
C 5	18.8	81.8	645	10 US-09-893-737-79	Sequence 79, Appli
6	17.2	78.2	528	9 US-09-738-626-1927	Sequence 1927, Ap
7	17.2	78.2	1969	10 US-09-864-761-3972	Sequence 3972, Ap
8	17.2	78.2	4575	9 US-10-240-198-1	Sequence 1, Appli
9	17.2	78.2	4575	10 US-09-782-378A-1	Sequence 1, Appli
10	17.2	78.2	4675	10 US-09-782-378A-2	Sequence 2, Appli
11	17.2	78.2	4679	9 US-10-038-972A-12	Sequence 12, Appli
12	17.2	78.2	4679	10 US-09-804-898-1	Sequence 1, Appli
13	17.2	78.2	4679	10 US-09-945-681-10	Sequence 10, Appli
14	17.2	78.2	4680	9 US-10-077-294-1	Sequence 1, Appli
15	17.2	78.2	4680	9 US-10-163-886-1	Sequence 1, Appli
16	17.2	78.2	4680	9 US-10-263-127-1	Sequence 1, Appli
17	17.2	78.2	8151	9 US-10-205-942-2	Sequence 2, Appli
18	17.2	78.2	3309400	9 US-09-738-626-1	Sequence 1, Appli
C 19	16.8	76.4	74868	9 US-10-175-523-67	Sequence 67, Appli

C 20	16.4	74.5	481	9 US-09-918-995-34768	Sequence 34768, A
C 21	16.4	74.5	550	10 US-09-864-761-9815	Sequence 9815, Ap
C 22	16.2	73.6	865	9 US-09-774-639-52	Sequence 52, Appli
C 23	16.2	73.6	865	9 US-09-969-730-89	Sequence 89, Appli
C 24	16.2	73.6	1575	10 US-09-815-242-7460	Sequence 7460, Ap
C 25	16.2	73.6	2673	10 US-09-764-877-3467	Sequence 3467, Ap
C 26	16.2	73.6	3538	10 US-09-764-877-3466	Sequence 3466, Ap
C 27	16.2	73.6	3656	9 US-09-764-872-824	Sequence 824, App
C 28	16.2	73.6	9551	9 US-09-764-872-825	Sequence 825, App
C 29	16.2	73.6	15558	9 US-09-764-872-823	Sequence 823, App
C 30	16.2	73.6	31124	9 US-10-060-763-12	Sequence 12, Appli
C 31	16.2	73.6	31124	12 US-10-063-763-12	Sequence 12, Appli
C 32	16	72.7	32220	10 US-09-764-877-3933	Sequence 3933, Ap
C 33	16	72.7	66479	9 US-10-041-856-1	Sequence 1, Appli
C 34	15.8	71.8	408	10 US-09-960-352-12444	Sequence 12444, A
C 35	15.8	71.8	422	9 US-09-918-995-12144	Sequence 12144, A
C 36	15.8	71.8	1516	10 US-09-745-605-2	Sequence 2, Appli
C 37	15.8	71.8	4398	9 US-10-105-989-1	Sequence 1, Appli
C 38	15.8	71.8	9927	10 US-09-070-927A-60	Sequence 60, Appli
C 39	15.8	71.8	85548	9 US-10-175-523-75	Sequence 75, Appli
C 40	15.8	71.8	1830121	9 US-10-329-960-1	Sequence 1, Appli
C 41	15.6	70.9	76	9 US-09-931-732-15	Sequence 15, Appli
C 42	15.6	70.9	76	9 US-10-142-566-6	Sequence 6, Appli
C 43	15.6	70.9	100	9 US-09-931-732-16	Sequence 16, Appli
C 44	15.6	70.9	103	9 US-10-142-566-7	Sequence 7, Appli
C 45	15.6	70.9	165	10 US-09-878-574-14461	Sequence 14461, A

#### ALIGNMENTS

RESULT 1  
US-10-017-273A-4/c  
; Sequence 4, Application US/10017273A  
; Publication No. US20030119714A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Inc.  
; APPLICANT: Naylor, Alasdair M.  
; APPLICANT: Van Der Graaf, Pieter H  
; APPLICANT: Wayman, Christopher P.  
; TITLE OF INVENTION: Treatment of Male Sexual Dysfunction  
; FILE REFERENCE: PC22013  
; CURRENT APPLICATION NUMBER: US/10/017,273A  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/265,358  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: GB 0030647.2  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: GB 0108730.3  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: GB 0120679.6  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 09/905,846  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/291,722  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: US 09/895,367  
; PRIOR FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 2893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-017-273A-4

Query Match 85.5%; Score 18.8; DB 9; Length 2893;  
Best Local Similarity 90.9%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGTTCATTCCTCCAGATGAAGAG 22  
|||||  
DB 908 GGTTCGTCGTTCCTCCAGATGAAGAG 887

## RESULT 6

US-09-738-626-1927  
; Sequence 1927, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1927  
; LENGTH: 528  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1927

Query Match 78.2%; Score 17.2; DB 9; Length 528;  
Best Local Similarity 86.4%; Pred. No. 81;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GGTATCATTCAGATGAG 22  
|||||  
Db 405 GGTATCATTCAGATGAG 426

RESULT 7  
US-09-864-761-3972  
; Sequence 3972, Application US/09864761  
; Patent No. US2002048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Acomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 3972  
; LENGTH: 1969  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005822.1  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13.  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 13  
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 14  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 12  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4  
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 12  
US-09-864-761-3972

Query Match 78.2%; Score 17.2; DB 10; Length 1969;  
Best Local Similarity 86.4%; Pred. No. 98;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GGTATCATTCAGATGAG 22  
|||||  
Db 1637 GGACACCATTCAGATGAG 1658

RESULT 8  
US-10-240-198-1  
; Sequence 1, Application US/10240198  
; Publication No. US20030100115A1  
; GENERAL INFORMATION:  
; APPLICANT: BTG International Ltd  
; APPLICANT: BEARD DR, PETER  
; APPLICANT: RAJ DR, KENNETH  
; TITLE OF INVENTION: CYTOTOXIC AGENTS  
; FILE REFERENCE: 142184W0  
; CURRENT APPLICATION NUMBER: US/10/240,198  
; CURRENT FILING DATE: 2002-09-30  
; PRIOR APPLICATION NUMBER: 0009887.1  
; PRIOR FILING DATE: 2000-04-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4675  
; TYPE: DNA  
; ORGANISM: adeno-associated virus 2  
US-10-240-198-1

Query Match 78.2%; Score 17.2; DB 9; Length 4675;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| ||| |||| |||||  
 Db 3865 GGTCATGATTACAGACGAAGAG 3886

RESULT 9  
 US-09-782-378A-1  
 ; Sequence 1, Application US/09782378A  
 ; Patent No. US20020102731A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hearing, Patrick  
 ; APPLICANT: Bahou, Wadie  
 ; APPLICANT: Sandalon, Ziv  
 ; APPLICANT: Gnatenko, Dmitri  
 ; TITLE OF INVENTION: Adenoviral Vectors  
 ; FILE REFERENCE: STONYB-04970  
 ; CURRENT APPLICATION NUMBER: US/09/782,378A  
 ; CURRENT FILING DATE: 2001-02-12  
 ; PRIOR APPLICATION NUMBER: 60/237,747  
 ; PRIOR FILING DATE: 2000-10-02  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 4675  
 ; TYPE: DNA  
 ; ORGANISM: Human adeno-associated virus 2  
 US-09-782-378A-1

Query Match 78.2%; Score 17.2; DB 10; Length 4675;  
 Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| ||| |||| |||||  
 Db 3865 GGTCATGATTACAGACGAAGAG 3886

RESULT 10  
 US-09-782-378A-2  
 ; Sequence 2, Application US/09782378A  
 ; Patent No. US20020102731A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hearing, Patrick  
 ; APPLICANT: Bahou, Wadie  
 ; APPLICANT: Sandalon, Ziv  
 ; APPLICANT: Gnatenko, Dmitri  
 ; TITLE OF INVENTION: Adenoviral Vectors  
 ; FILE REFERENCE: STONYB-04970  
 ; CURRENT APPLICATION NUMBER: US/09/782,378A  
 ; CURRENT FILING DATE: 2001-02-12  
 ; PRIOR APPLICATION NUMBER: 60/237,747  
 ; PRIOR FILING DATE: 2000-10-02  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 4675  
 ; TYPE: DNA  
 ; ORGANISM: Human adeno-associated virus 2  
 US-09-782-378A-2

Query Match 78.2%; Score 17.2; DB 10; Length 4675;  
 Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| ||| |||| |||||  
 Db 3865 GGTCATGATTACAGACGAAGAG 3886

RESULT 11  
 US-10-038-972A-12  
 ; Sequence 12, Application US/10038972A  
 ; Publication No. US20020192823A1

; GENERAL INFORMATION:  
 ; APPLICANT: J. Bartlett  
 ; TITLE OF INVENTION: RAV VECTORS AND METHODS  
 ; FILE REFERENCE: 28335/36996US  
 ; CURRENT APPLICATION NUMBER: US/10/038,972A  
 ; CURRENT FILING DATE: 2002-01-04  
 ; PRIOR APPLICATION NUMBER: US 60/260,124  
 ; PRIOR FILING DATE: 2001-01-05  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 12  
 ; LENGTH: 4679  
 ; TYPE: DNA  
 ; ORGANISM: adeno-associated virus 2  
 US-10-038-972A-12

Query Match 78.2%; Score 17.2; DB 9; Length 4679;  
 Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| ||| |||| |||||  
 Db 3870 GGTCATGATTACAGACGAAGAG 3891

RESULT 12  
 US-09-804-898-1  
 ; Sequence 1, Application US/09804898  
 ; Patent No. US20020045264A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DURING, MATTHEW  
 ; APPLICANT: XIAO, WEIDONG  
 ; TITLE OF INVENTION: PRODUCTION OF CHIMERIC CAPSID VECTORS  
 ; FILE REFERENCE: 102182-14  
 ; CURRENT APPLICATION NUMBER: US/09/804,898  
 ; CURRENT FILING DATE: 2001-03-13  
 ; PRIOR APPLICATION NUMBER: 60/189,110  
 ; PRIOR FILING DATE: 2000-03-14  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 4679  
 ; TYPE: DNA  
 ; ORGANISM: adeno-associated virus 2  
 US-09-804-898-1

Query Match 78.2%; Score 17.2; DB 10; Length 4679;  
 Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGTCATCATTCAGATGAAGAG 22  
 ||||| ||| |||| |||||  
 Db 3870 GGTCATGATTACAGACGAAGAG 3891

RESULT 13  
 US-09-945-681-10  
 ; Sequence 10, Application US/09945681  
 ; Patent No. US20020064878A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UNIVERSITE DE NANTES  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR RECOMBINANT  
 ; FILE OF INVENTION: ADENO-ASSOCIATED VIRUS PRODUCTION  
 ; FILE REFERENCE: B4182AA - UNIVERSITE DE NANTES  
 ; CURRENT APPLICATION NUMBER: US/09/945,681  
 ; CURRENT FILING DATE: 2001-09-05  
 ; PRIOR APPLICATION NUMBER: PCT/EP 00/01854  
 ; PRIOR FILING DATE: 2000-03-05  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 10  
 ; LENGTH: 4679  
 ; TYPE: DNA

ORGANISM: adeno-associated virus 2  
US-09-945-681-10

Query Match 78.2%; Score 17.2; DB 10; Length 4679;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GGTCATCATTCACGATGAAGAG 22  
||||| ||| ||| ||| ||| ||| |||  
Db 3870 GGTCATGATTACAGCAGAG 3891

RESULT 14

US-10-077-294-1  
; Sequence 1, Application US/10077294  
; Patent No. US20020159979A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.  
; TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/077.294  
FILING DATE: 15-Feb-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/691.604  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: No. US20020159979Aland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-077-294-1  
Query Match 78.2%; Score 17.2; DB 9; Length 4680;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GGTCATCATTCACGATGAAGAG 22  
||||| ||| ||| ||| ||| ||| |||  
Db 3870 GGTCATGATTACAGCAGAG 3891

RESULT 15  
US-10-163-886-1  
; Sequence 1, Application US/10163886  
; Publication No. US20020187129A1  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Philip R.

Query Match 78.2%; Score 17.2; DB 9; Length 4680;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GGTCATCATTCACGATGAAGAG 22  
||||| ||| ||| ||| ||| ||| |||  
Db 3870 GGTCATGATTACAGCAGAG 3891

TITLE OF INVENTION: Adeno-Associated Virus Materials and Methods

NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/163,886  
FILING DATE: 04-Jun-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/292,703  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: No. US20020187129Aland, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31975  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4680 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-163-886-1  
Query Match 78.2%; Score 17.2; DB 9; Length 4680;  
Best Local Similarity 86.4%; Pred. No. 1.1e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GGTCATCATTCACGATGAAGAG 22  
||||| ||| ||| ||| ||| ||| |||  
Db 3870 GGTCATGATTACAGCAGAG 3891

Search completed: July 9, 2003, 02:22:08  
Job time : 130.113 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: July 8, 2003, 00:47:28 ; Search time 1115.35 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-12

Perfect score: 22  
Sequence: 1 ggctcatccagatgaagag 22

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estpi: \*  
7: em\_estro: \*  
8: em\_estc: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_estc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vit: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rod: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	20.4	92.7	600	13	BI989738 4044-75 M
c 2	19	86.4	388	9	AI978436 R2565, R E
c 3	18.8	85.5	682	17	AZ335729 LM0065L20
c 4	18.4	83.6	436	9	AA644490 af73401.r
c 5	18.4	83.6	605	17	AQ568304 HS_5237_B
c 6	18.4	83.6	740	12	BG035328 602324955

c	7	18	81.8	387	12	BF944641
c	8	17.8	80.9	520	9	AI558754
c	9	17.8	80.9	586	12	BF186831
c	10	17.8	80.9	587	10	BE455509
c	11	17.8	80.9	637	9	AI055512
c	12	17.8	80.9	642	9	AA739656
c	13	17.8	80.9	689	13	BM110208
c	14	17.8	80.9	766	14	BQ118381
c	15	17.8	80.9	869	12	BQ476576
c	16	17.4	79.1	347	10	AW702998
c	17	17.4	79.1	354	12	BG406112
c	18	17.4	79.1	377	10	AW757292
c	19	17.4	79.1	384	14	BM953996
c	20	17.4	79.1	404	12	BE801447
c	21	17.4	79.1	423	9	AI495501
c	22	17.4	79.1	462	10	BE209730
c	23	17.4	79.1	488	14	BM86968
c	24	17.4	79.1	494	17	AQ140402
c	25	17.4	79.1	508	9	AU060271
c	26	17.4	79.1	532	17	AQ675598
c	27	17.4	79.1	559	17	AQ459342
c	28	17.4	79.1	566	13	BQ325350
c	29	17.4	79.1	602	13	BQ325377
c	30	17.4	79.1	659	9	AV247976
c	31	17.4	79.1	695	17	BH485426
c	32	17.4	79.1	878	12	BF275584
c	33	17.4	79.1	915	17	CNS03081
c	34	17.2	78.2	164	10	AW935060
c	35	17.2	78.2	396	14	BQ320133
c	36	17.2	78.2	396	14	BQ320555
c	37	17.2	78.2	435	10	AW752398
c	38	17.2	78.2	506	12	BG885559
c	39	17.2	78.2	525	10	BE400968
c	40	17.2	78.2	550	17	BG2006
c	41	17.2	78.2	573	17	AQ726470
c	42	17.2	78.2	582	13	BQ457294
c	43	17.2	78.2	591	17	AZ798018
c	44	17.2	78.2	594	10	BE583834
c	45	17.2	78.2	600	12	BG020270

#### ALIGNMENTS

RESULT 1  
BI989738/c  
LOCUS 4044-75 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, EST 20-DEC-2001  
DEFINITION mRNA sequence.  
ACCESSION BI989738  
VERSION BI989738.1 GI:17960740  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 600)  
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W., White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

FEATURES  
Source  
Location/Qualifiers  
1..600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

```

from etiolated leaf sheath from the cultivar IR36
(O.sativa). "
BASE COUNT      129 a      82 c      84 g      87 t      6 others
ORIGIN

Query Match      86.4%; Score 19; DB 9; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2  GTCATCATTCAGATGAAG 20
          |||||
Db      254 GTCATCATTCAGATGAAG 272

```

RESULT	3
AZ335729/c	
LOCUS	
DEFINITION	AZ335729 linear DNA 682 bp GSS 29-SEP-2000 M0065L20R Mouse 10kb plasmid UUGClm library Mus musculus genomic clone UUGClm0065L20 R, DNA sequence.
ACCESSION	AZ335729
VERSION	AZ335729.1 GI:10404330

SOURCE  
house mouse.  
Mus musculus

ORGANISM  
Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;  
Eukaryota; Metazoa; Mammalia; Eutheria; Chordata; Vertebrata; Euteleostomi;  
1 (bases 1 to 692)

REFERENCE  
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.,  
and Wright, D., Weiss, R.

TITLE  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

CONTACT: ROBERT B. WEISS  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: [ddunne@genetics.utah.edu](mailto:ddunne@genetics.utah.edu)  
Insert Length: 10000 Std Error: 0.00  
Plate: 0065 row: L column: 20  
Seq primer: CACACAGGAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 682.  
Location/Qualifiers  
1. .682  
/organism="Mus musculus"  
/strain="C57BL/6J"

/clone.lib="Mouse\_10kb plasmid UUGc1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /note="Vector: pWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (<http://www.jax.org/resources/documents/dnares/>). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi14732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 740)  
 NTH-MGC http://mgc.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LAM10138 row: n column: 14  
 High quality sequence stop: 569.  
 Location/Qualifiers

## FEATURES

source

1. 740  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="4413133"  
 /clone\_lib="NTH\_MGC\_90"  
 /tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH\_MGC Library."  
 180 a 169 c 189 g 202 t

## FEATURES

source

Query Match 83.6%; Score 18.4; DB 12; Length 740;  
 Best Local Similarity 95.0%; Pred. No. 3.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 TCATCATCCAGTGAAGAG 22  
 |||||  
 Db 379 TCATCATCCAGTGAAGAG 360  
 BASE COUNT 180 a 169 c 189 g 202 t  
 ORIGIN

Query Match 83.6%; Score 18.4; DB 12; Length 740;  
 Best Local Similarity 95.0%; Pred. No. 3.6e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 TCATCATCCAGTGAAGAG 22  
 |||||  
 Db 379 TCATCATCCAGTGAAGAG 360  
 BASE COUNT 180 a 169 c 189 g 202 t  
 ORIGIN

RESULT 7  
 BF944641/c  
 LOCUS  
 DEFINITION PM0-N1171-181000-001-h12 N1171 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF944641  
 VERSION BF944641.1 GI:12361916  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 387)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

RESULT 7  
 BF944641/c  
 LOCUS  
 DEFINITION PM0-N1171-181000-001-h12 N1171 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF944641  
 VERSION BF944641.1 GI:12361916  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 387)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl=PM0-N1171-181000-001-h12&t3=2000-10-18&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 387.  
 Location/Qualifiers

## FEATURES

source

1. 387  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="N1171"  
 /dev\_stage="Adult"  
 /note="Organ: nervous\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
 101 a 75 c 94 g 117 t

Query Match 81.8%; Score 18; DB 12; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 ATCATCCAGTGAAGAG 22  
 |||||  
 Db 155 ATCATCCAGTGAAGAG 138  
 BASE COUNT 101 a 75 c 94 g 117 t  
 ORIGIN

Query Match 81.8%; Score 18; DB 12; Length 387;  
 Best Local Similarity 100.0%; Pred. No. 4.5e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 ATCATCCAGTGAAGAG 22  
 |||||  
 Db 155 ATCATCCAGTGAAGAG 138  
 BASE COUNT 101 a 75 c 94 g 117 t  
 ORIGIN

RESULT 8  
 A1558754  
 LOCUS  
 DEFINITION f80b01.y1 zebrafish WashU MPIMG EST Danio rerio cDNA clone  
 IMAGE:3718153 5', mRNA sequence.  
 ACCESSION A1558754  
 VERSION A1558754.1 GI:4508992  
 KEYWORDS EST  
 SOURCE zebrafish.  
 ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 520)  
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
 WashU Zebrafish EST Project 1998  
 Unpublished (1998)  
 Contact: Stephen L. Johnson  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: zbrfish@wustl.edu

CNA Library Preparation: Matthew Clark. cDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center. Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address: www.rzpd.de)  
 Seq primer: T3 ET from Amersham  
 High quality sequence stop: 489.  
 POLYA-No.  
 Location/Qualifiers

## FEATURES

source

1. 520  
 /organism="Danio rerio"

```

/db_xref="taxon:7955"
/clone="IMAGE:3718153"
/clone_lib="Zebrafish Washu MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/note="vector: pSPOR1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'-pGACTAGTCTGAGATCGAGCGCCGCTTTTCTTTTCTTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPOR1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
BASE COUNT      159 a      106 c      129 g      126 t
ORIGIN
Query Match      80.9%; Score 17.8; DB 9; Length 520;
Best Local Similarity 90.5%; Pred. No. 6.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      2  GTCATCATCCAGATGAAGAG 22
||||| ||||| ||||| ||||| |||||
Db      428  GTCACCATCCATATGAAGAG 448

RESULT 9
BF186831
LOCUS      BF186831      586 bp      mRNA      linear      EST 01-NOV-2000
DEFINITION      E443118 potato stolon, Cornell University Solanum tuberosum cDNA.
clone cSTA37C3 5' sequence, mRNA sequence.
ACCESSION      BF186831
VERSION        BF186831.1 GI:11069050
KEYWORDS       potato.
SOURCE         Solanum tuberosum
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE      1 (bases 1 to 586)
AUTHORS       van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R.,
Holt,I.E., Liang,F., Hansen,T.S., Utterback,T., Bowman,C.L., Doan
,B., Bougri,O., Buell,C.R., Ronning,C.M., Tanksley,S.D. and Baker
,B.
TITLE         Generation of ESTs from potato swelling stolons
JOURNAL        Unpublished (1999)
COMMENT       Contact: TIGR
The Institute for Genomic Research
Rockville, MD 20850, USA
Tel: (301)838-0200
Email: http://www.tigr.org/tdb/potato/
For clone information, please contact Research Genetics, Libraries
Division, tel# 1-800-711-6195, email- cdna@resgen.com.
FEATURES             Location/Qualifiers
     source          1..586
                     /organism="Solanum tuberosum"
                     /cultivar="Bintje"
                     /db_xref="taxon:4113"
                     /clone="cSTA37C3"
                     /clone_lib="potato.stolon, Cornell University"
                     /tissue_type="axillary buds of stem explants, swelling
                     stolons"

/dev_stage="1 to 3 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; RNA was supplied by Christian Bachem & Beatrice
Horvath(Laboratory of Plant Breeding, Dept. of Plant
Sciences, Wageningen University, The Netherlands). Total
RNA was isolated from developing axillary buds of potato
nodal stem cuttings cultured on medium for the
introduction of tuber formation as described in Bachem et
al. (Plant Journal 1996). Tissue samples were taken of
stages corresponding to growing stolons and the early
stages of tuber formation."
BASE COUNT      193 a      117 c      123 g      153 t
ORIGIN
Query Match      80.9%; Score 17.8; DB 12; Length 586;
Best Local Similarity 90.5%; Pred. No. 6.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy      2  GTCATCATCCAGATGAAGAG 22
||||| ||||| ||||| ||||| |||||
Db      337  GTTATCATCTCTGTGAAGAG 357

RESULT 10
BE455509
LOCUS      BE455509      587 bp      mRNA      linear      EST 22-OCT-2001
DEFINITION      HVSMEG0017115f Hordeum vulgare pre-anthesis spike EST library
HVCNDA0008 (white to yellow anther) Hordeum vulgare cDNA clone
HVSMEG0017115f, mRNA sequence.
ACCESSION      BE455509
VERSION        BE455509.2 GI:13154820
KEYWORDS       EST.
SOURCE         Hordeum vulgare.
ORGANISM       Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE      1 (bases 1 to 587)
AUTHORS       Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Chol,D.W., Fenton
,R.D., Close,S.J., Oates,R. and Main,D.
TITLE         Development of a genetically and physically anchored EST resource
for barley genomics: Morex pre-anthesis spike cDNA library
JOURNAL        Unpublished (2001)
COMMENT       On Jul 26, 2000 this sequence version replaced gi:9465256.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 200
Seq primer: ATTTACCTCTACTAAGGG
High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     source          1..587
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEG0017115f"
                     /clone_lib="Hordeum vulgare pre-anthesis spike EST library
                     HVCNDA0008 (white to yellow anther)"
                     /tissue_type="pre-anthesis spike"
                     /lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
Plants were grown in the greenhouse at the University of
California, Riverside (Fenton, SJ Close, TJ Close). Whole
spike with awns trimmed were collected at white, green and
yellow anther stages (Fenton). Total RNA was prepared from
each pool, equal quantities of all three RNA pools were
combined, poly(A) RNA was purified from the mixture, one

```

BASE COUNT	161 a	125 c	140 g	211 t
ORIGIN				
Query Match		80.9%	Score 17.8;	DB 9;
Best Local Similarity		90.5%	Pred. No. 6.6e+02;	Length 637;

ORGANISM	Source	Potato
<i>Solanum tuberosum</i>		
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanaceae; Solanum.		

REFERENCE 1 (bases 1 to 689)  
AUTHORS van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,  
Utterback,T., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C.,  
Tanksley,S. and Baker,B.  
TITLE Generation of ESTs from potato roots  
JOURNAL Unpublished (2001)  
COMMENT Contact: Research Genetics, Libraries Division  
Tel: 1-800-711-6195  
Email: cda@resgen.com  
For clone info: please contact Research Genetics, Libraries  
Division tel 1-800-711-6195, email cda@resgen.com  
Seq primer: T3.  
FEATURES Location/Qualifiers  
source 1..689  
/organism="Solanum tuberosum"  
/cultivar="Kennebec"  
/db\_xref="taxon:4113"  
/clone="cPRO7M1"  
/clone\_lib="potato roots"  
/tissue\_type="roots"  
/dev\_stage="in vitro grown stem cuttings"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Cornell University, Tanksley lab;  
sequencing: The Institute for Genomic Research. Roots were  
isolated from in vitro grown stem cuttings on CM medium.  
Roots were isolated two weeks after placing the stem  
cuttings from in vitro grown plants on medium."  
BASE COUNT 219 a 142 c 137 g 191 t  
ORIGIN  
Query Match 80.9%; Score 17.8; DB 13; Length 689;  
Best Local Similarity 90.5%; Pred. No. 6.8e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 2  
QY 2 GTCATCATTCAGATGAAGAG 22  
DB 349 GTTATCATTCCTGATGAAGAG 369  
RESULT 14  
LOCUS BQ118381 766 bp mRNA linear EST 17-APR-2002  
DEFINITION EST603957 mixed potato tissues Solanum tuberosum cDNA clone STMEA47  
3' end, mRNA sequence.  
ACCESSION BQ118381  
VERSION BQ118381.1 GI:20170343  
KEYWORDS EST.  
SOURCE potato.  
ORGANISM Solanum tuberosum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
REFERENCE 1 (bases 1 to 766)  
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,  
Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and  
Karamycheva,S.A.  
TITLE Generation of a set of potato cDNA clones for microarray analyses  
JOURNAL Unpublished (2002)  
COMMENT Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato@tigr.org  
This clone is available through the Research Genetics, contact the  
Research Genetics for further information 1-800-711-6195 or  
cda@resgen.com  
Seq primer: T7.  
FEATURES Location/Qualifiers  
source 1..766  
/organism="Solanum tuberosum"  
/cultivar="Kennebec or Binjete"  
/db\_xref="taxon:4113"  
/clone="STMEA47"

/clone\_lib="mixed potato tissues"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/note="Vector: pBluescript SK(-); Site\_1: EcoRI; Site\_2:  
XhoI; supplier: Combination of untreated and Phytophthora  
infestans-treated libraries of stolons, leaves, leaflets,  
axillary buds of stem explants, petioles, germinating eyes  
, tubers, or roots."  
BASE COUNT 224 a 151 c 150 g 241 t  
ORIGIN  
Query Match 80.9%; Score 17.8; DB 14; Length 766;  
Best Local Similarity 90.5%; Pred. No. 7.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 GTCATCATTCAGATGAAGAG 22  
DB 428 GTTATCATTCCTGATGAAGAG 408  
RESULT 15  
LOCUS BG476576 869 bp mRNA linear EST 21-MAR-2001  
DEFINITION 602524536F1 NIH\_MGC\_20 Homo sapiens cDNA clone IMAGE:4642976 5'  
mRNA sequence.  
ACCESSION BG476576  
VERSION BG476576.1 GI:13408855  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 869)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
plate: L1CM1412 row: g column: 09  
High quality sequence stop: 749.  
FEATURES Location/Qualifiers  
source 1..869  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4642976"  
/clone\_lib="NIH\_MGC\_20"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pORF7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 234 a 231 c 225 g 179 t  
ORIGIN  
Query Match 80.9%; Score 17.8; DB 12; Length 869;  
Best Local Similarity 90.5%; Pred. No. 7.4e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GGTATCATTCAGATGAAGA 21  
DB 820 GGACACATTCAGATGAAGA 840

Search completed: July 8, 2003, 09:21:43  
Job time : 1120.5 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 220.098 Seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-13

Perfect score: 20

Sequence: 1 cgaatgagcgcgcctgttg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_ba:\*
  - 2: gb\_hg:\*
  - 3: gb\_in:\*
  - 4: gb\_om:\*
  - 5: gb\_ov:\*
  - 6: gb\_pat:\*
  - 7: gb\_ph:\*
  - 8: gb\_pl:\*
  - 9: gb\_pr:\*
  - 10: gb\_ro:\*
  - 11: gb\_sts:\*
  - 12: gb\_sy:\*
  - 13: gb\_un:\*
  - 14: gb\_vi:\*
  - 15: em\_ba:\*
  - 16: em\_fun:\*
  - 17: em\_hum:\*
  - 18: em\_in:\*
  - 19: em\_mu:\*
  - 20: em\_om:\*
  - 21: em\_or:\*
  - 22: em\_ov:\*
  - 23: em\_pat:\*
  - 24: em\_ph:\*
  - 25: em\_pl:\*
  - 26: em\_ro:\*
  - 27: em\_sts:\*
  - 28: em\_un:\*
  - 29: em\_vi:\*
  - 30: em\_htg\_hum:\*
  - 31: em\_htg\_inv:\*
  - 32: em\_htg\_other:\*
  - 33: em\_htg\_mus:\*
  - 34: em\_htg\_pln:\*
  - 35: em\_htg\_rod:\*
  - 36: em\_htg\_mam:\*
  - 37: em\_htg\_vrt:\*
  - 38: em\_sy:\*
  - 39: em\_htgo\_hum:\*
  - 40: em\_htgo\_mus:\*
  - 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	20	100.0	20	6	AX014713	Sequence
2	20	100.0	2076	6	AX146976	Sequence
3	20	100.0	2232	6	AX139864	Sequence
4	20	100.0	2262	6	AX146980	Sequence
5	20	100.0	2340	6	AX146978	Sequence
6	20	100.0	2340	6	AX473102	Sequence
7	20	100.0	2636	6	AX139743	Sequence
8	20	100.0	2663	6	AX139745	Sequence
9	20	100.0	2676	6	AX033274	Sequence
10	20	100.0	2714	6	AX139741	Sequence
11	20	100.0	2765	6	AX014701	Sequence
12	20	100.0	2784	9	AF336981	Homo sapi
13	20	100.0	2850	9	AK093058	Sequence
14	20	100.0	2893	6	AX356951	Sequence
15	20	100.0	2893	6	AX463057	Sequence
16	20	100.0	2953	6	AX473100	Sequence
17	20	100.0	2975	6	AX356955	Sequence
18	20	100.0	2975	6	AX463058	Sequence
19	20	100.0	25807	9	AL589746	Human DNA
20	20	100.0	174953	2	AC094732	Rattus no
21	18.4	92.0	280	11	G70475	721684831FB
22	18.4	92.0	288	11	G71420	G71389 716249731FM
23	18.4	92.0	316	11	G71389	AF407018 Escherich
24	18.4	92.0	720	1	AF407018	AF302075 Mus muscu
25	18.4	92.0	2583	10	AF302075	AF157106 Mus muscu
26	18.4	92.0	2601	10	AF157106	AF302076 Mus muscu
27	18.4	92.0	2652	10	AF302076	AF302077 Mus muscu
28	18.4	92.0	2694	10	AF302077	M94104 Escherichia
29	18.4	92.0	2886	1	ECOPROT5	AF157105 Mus muscu
30	18.4	92.0	2892	10	AF157105	AX033272 Sequence
31	18.4	92.0	2925	6	AX033272	AF176569 Mus muscu
32	18.4	92.0	2925	10	AF176569	AE000446 Escherich
33	18.4	92.0	10120	1	AE000446	AE005600 Escherich
34	18.4	92.0	12642	1	ECOUW82	L10328 E. coli; th
35	18.4	92.0	136254	1	ECOUW82	AL607032 Mus muscu
36	18.4	92.0	208249	2	AL607032	AC023611 Mus muscu
37	18.4	92.0	221285	2	AC023611	AC073670 Mus muscu
38	18.4	92.0	252977	2	AC073670	AC020870 Mus muscu
39	18.4	92.0	256373	2	AC020870	AC090533 Mus muscu
40	18.4	92.0	272545	2	AC090533	AP002566 Escherich
41	18.4	92.0	280900	1	AP002566	L11256 Serratia ma
42	16.8	84.0	579	1	SWAINFCA	X63841 A.caulinoda
43	16.8	84.0	6881	1	ACNTRYXA	AE004496 Pseudomon
44	16.8	84.0	10520	1	AE004496	AE013793 Yersinia
45	16.8	84.0	11127	1	AE013793	

ALIGNMENTS

RESULT 1	AX014713	Sequence 13 from Patent WO9953077.	20 bp	DNA	linear	PAT 07-SEP-2000
LOCUS	AX014713					
DEFINITION	Sequence 13 from Patent WO9953077.					
ACCESSION	AX014713					
VERSION	AX014713.1	GI:10040986				
KEYWORDS						
SOURCE		synthetic construct.				
ORGANISM		synthetic construct				
REFERENCE		artificial sequences.				
AUTHORS		1 (bases 1 to 20)				
		Schwartz J.C., Gros C., Oulmet T., Rose C., Bonhomme M.C. and				
		Facchinetti P.				
TITLE		Novel nep ii membrane metalloprotease and its use for screening				
		inhibitors useful in therapy				

JOURNAL Patent: WO 9953077-A 13 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES  
source  
1. .20  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
3 a 5 c 8 g 4 t  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGAGCGCGCTGTTG 20  
|||||  
Db 1 CGATGAGGAGCGCGCTGTTG 20  
|||||  
RESULT 2  
AX146976/c  
LOCUS 2076 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 1 from Patent WO0136610.  
ACCESSION AX146976  
VERSION AX146976.1 GI:14346247  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2076)  
AUTHORS Deleersnijder,W., Wiegers,R. and Weske,M.  
TITLE Human enzymes of the metalloprotease family  
JOURNAL Patent: WO 0136610-A.1 25-MAY-2001;  
Solvay Pharmaceuticals B.V. (NL)  
FEATURES  
source  
1. .2076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS  
1. .2076  
/note="unnamed protein product"  
/protein\_start=1  
/codon\_start=1  
/db\_xref="GI:14346248"  
/translation="CTTTCVIAAARILQNMPTTEPCDDFYQFACGGWLRHVIPET  
NSRYIFDLRDELEVLKAVLENSTAKDRPAVERKARTLYRSCMNSQSVTEKRGSOPLL  
DILEVVGWPVAMDRWNETVGLWELEKALMNSQFNRRVLIDLFINWDDONSRRHI  
LYIDOPTLGMPSREYFNGSGNRKYREAYLQFVSVATILLREDANLPDSCLVQEDMM  
QVLETLQAKATVPOEERHDVIALYHRMGLLEQLSQFGLKGNWTLFTOTVLSSVKI  
KLLPDEEVVYIGIPYQLQNLNIIDYTSARTIQNYLVRLVLDRIQSLQRFKDRVNY  
RKALFTVVEVYWRCEGVYVNSNMENAGSLYVREAFPGDSKSMVRELIDKRVTVFV  
ETLDELGWDESKKAOKKAMSIREQIGHDPYILEEMNRRLDEEYSNLFSEDIYFE  
NSLONLKYGAOBSLRKLEKVDPNLWIGAAVYNAFYSPNRQIVFPAGILOPPEFSK  
EQPOLNFGGIGWVGHETHGFDNGRNFNGMNMWNSFQHFREQSECHIYQ  
YGNYSWDLADEQNVNGFNLGNIADNGVRQAYKAYLKWMAEGGKQDQLPLDTHHE  
QFFNTYAQVWCGSYRPEFAIQSIKTDVHSPLKRYVLGSLQNLAAAFADTFHCARGTTP  
HPKERCVRW"  
BASE COUNT 493 a 578 c 636 g 369 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 2076;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGAGCGCGCTGTTG 20  
|||||  
Db 436 CGATGAGGAGCGCGCTGTTG 417  
|||||  
RESULT 3  
AX19864/c

LOCUS AX319864 2232 bp DNA linear PAT 14-DEC-2001  
DEFINITION Sequence 28 from Patent WO0183782.  
ACCESSION AX319864  
VERSION AX319864.1 GI:17901454  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Plozman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and  
Payne,V.  
TITLE Novel proteases  
JOURNAL Patent: WO 0183782-A 28 NOV-2001;  
Sugen, Inc. (US)  
FEATURES  
source  
1. .2232  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 512 a 620 c 705 g 395 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 2232;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGAGCGCGCTGTTG 20  
|||||  
Db 595 CGATGAGGAGCGCGCTGTTG 576  
|||||  
RESULT 4  
AX146980/c  
LOCUS 2262 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 5 from Patent WO0136610.  
ACCESSION AX146980  
VERSION AX146980.1 GI:14346251  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2262)  
AUTHORS Deleersnijder,W., Wiegers,R. and Weske,M.  
TITLE Human enzymes of the metalloprotease family  
JOURNAL Patent: WO 0136610-A 5 25-MAY-2001;  
Solvay Pharmaceuticals B.V. (NL)  
FEATURES  
source  
1. .2262  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS  
1. .2262  
/note="unnamed protein product"  
/protein\_start=1  
/codon\_start=1  
/db\_xref="GI:14346252"  
/translation="MGKSEGPVGMVESAGAGQKRPFGLEGLLLLVLTAALVALG  
VLYADRRGPEAQEVSEVCTPGCVIAAARILQNMPTTEPCDDFYQFACGGWLRHV  
IPETNSRYIFDLRDELEVLKAVLENSTAKDRPAVERKARTLYRSCMNSQSVTEKRG  
SOPLLDELVVGWPVAMDRWNETVGLWELEKALMNSQFNRRVLIDLFINWDDONS  
SRHIYIDOPTLGMPSREYFNGSGNRKYREAYLQFVSVATILLREDANLPDSCLVQ  
EDMMQVLETLQAKATVPOEERHDVIALYHRMGLLEQLSQFGLKGNWTLFTOTVL  
SVKIKLLPDEEVVYIGIPYQLQNLNIIDYTSARTIQNYLVRLVLDRIQSLQRFKDR  
VNYRKALFTVVEVYWRCEGVYVNSNMENAGSLYVREAFPGDSKSMVRELIDKRV  
TYFETLDELGWDESKKAOKKAMSIREQIGHDPYILEEMNRRLDEEYSNLFSEDI  
LYFENSLONLKYGAOBSLRKLEKVDPNLWIGAAVYNAFYSPNRQIVFPAGILOP  
PFSEKQPOLNFGGIGWVGHETHGFDNGRNFNGMNMWNSFQHFREQSECHIYQ  
MIYQYGNYSWDLADEQNVNGFNLGNIADNGVRQAYKAYLKWMAEGGKQDQLPLD  
LTHEQLEFFINAYQVWCGSYRPEFAIQSIKTDVHSPLKRYVLGSLQNLAAAFADTFHCAR  
GTPMHPKERCVRW"  
BASE COUNT 520 a 628 c 716 g 398 t  
ORIGIN

```

Query Match      100.0%; Score 20; DB 6; Length 2262;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
    |||||
Db 622 CGATGAGGACGCGCTGTTG 603

RESULT 5
AX146978/c
LOCUS AX146978 2340 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 3 from Patent WO0136610.
ACCESSION AX146978
VERSION AX146978.1 GI:14346249
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2340)
AUTHORS Deleersnijder, W., Wiegers, R. and Weske, M.
TITLE Human enzymes of the metalloprotease family
JOURNAL Patent: WO 0136610-A 3 25-MAY-2001;
Solvay Pharmaceuticals B.V. (NL)
FEATURES
    source
        Location/Qualifiers
            1..2340
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                <1..2340
                    /note="unnamed protein product"
                    /codon_start=1
                    /protein_id="CAC41160.1"
                    /db_xref="GI:14346250"
                    /translation="MGKSEGPVWVESAGRAGOKRPGFLEGGLLILLVTAALVALG
                    VLYADRRGKOLPLASRLCFLOERTEVKRPGIPEAQEVSEVCTTGGVIAAARIL
                    QNDPTPECDDEYQFACGGLRHVTPETNSYSIFDLRLDEVLKAVLENSTAK
                    DRPAVEKARTLYSCMQSVIERKGSQPLDILEVGGVPMVDRWNETVGLWELER
                    QALMNSQNRRLIDLFITWDDQNSRHIIYIDQPTLGMPSREYFNGGSKVREA
                    YLQFMYSVATLLREDANLPDSCVLQSDMMQVLELTQLAKATVPOERHDVIALYHR
                    MGLEEQSQGLGKFNWTLFIQTIVLSVKIKLLPDEEVVVGIPYLNLENIDITYSA
                    RTQNYLWRLVLDIGLSQREFKDRVNYRKALFGTMVEVRWRECVCYVYNSMENA
                    VGSLYVREAPPGSKSMVRELIDKRVFVETLDELGWMBDEESKKAQEKAMSIREQI
                    GHFDYILEENRRLDEYSINLFSEDLFENSQNLKVGAKRSKLRKRVDPNLMWII
                    GAAVNAFNSNQIVFPAGILQPPFSEKQYQYGNYSWDLADEQNVNGFNTLGTADNG
                    NFNGNMMDWNSFSTQHFREQSECMYQYGNYSWDLADEQNVNGFNTLGTADNG
                    GVROAYKAYLKWMAEGKQDQLPGLDITHEQLFFINTAQAQWCGSYRPEFAIOSIKTDV
                    HSPLKYRVLSLQNLAAAFATFHCACTGPMHPRCVRW"
BASE COUNT 539 a 649 c 739 g 413 t
ORIGIN

Query Match      100.0%; Score 20; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
    |||||
Db 700 CGATGAGGACGCGCTGTTG 681

RESULT 6
AX473102/c
LOCUS AX473102 2340 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 3 from Patent WO0226958.
ACCESSION AX473102
VERSION AX473102.1 GI:22207836
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Bandaru, R. and Silos-Santiago, I.

```

```

TITLE Human neprilysin protease
JOURNAL Patent: WO 0226958-A 3 04-APR-2002;
MILLENNIUM PHARMACEUTICALS, INC. (US)
FEATURES
    source
        Location/Qualifiers
            1..2340
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
BASE COUNT 538 a 647 c 740 g 415 t
ORIGIN

Query Match      100.0%; Score 20; DB 6; Length 2340;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
    |||||
Db 700 CGATGAGGACGCGCTGTTG 681

RESULT 7
AX139743/c
LOCUS AX139743 2636 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 3 from Patent EP1069188.
ACCESSION AX139743
VERSION AX139743.1 GI:14275325
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2636)
AUTHORS Jagerschmidt, A., Agnel, M. and Culouscou, J.M.
TITLE Three neprilysin-like membrane metalloproteinases
JOURNAL Patent: EP 1069188-A 3 17-JAN-2001;
SANOFTI-SYNTHELABO (FR)
FEATURES
    source
        Location/Qualifiers
            1..2636
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
            1..7
                /note="unnamed protein product"
            8..2269
                /codon_start=1
                /protein_id="CAC39975.1"
                /db_xref="GI:14275326"
                /translation="MGKSEGPVWVESAGRAGOKRPGFLEGGLLILLVTAALVALG
                VLYADRRGKOLPLASRLCFLOERTEVKRPGIPEAQEVSEVCTTGGVIAAARIL
                QNDPTPECDDEYQFACGGLRHVTPETNSYSIFDLRLDEVLKAVLENSTAK
                DRPAVEKARTLYSCMQSVIERKGSQPLDILEVGGVPMVDRWNETVGLWELER
                QALMNSQNRRLIDLFITWDDQNSRHIIYIDQPTLGMPSREYFNGGSKVREA
                YLQFMYSVATLLREDANLPDSCVLQSDMMQVLELTQLAKATVPOERHDVIALYHR
                MGLEEQSQGLGKFNWTLFIQTIVLSVKIKLLPDEEVVVGIPYLNLENIDITYSA
                RTQNYLWRLVLDIGLSQREFKDRVNYRKALFGTMVEVRWRECVCYVYNSMENA
                VGSLYVREAPPGSKSMVRELIDKRVFVETLDELGWMBDEESKKAQEKAMSIREQI
                GHFDYILEENRRLDEYSINLFSEDLFENSQNLKVGAKRSKLRKRVDPNLMWII
                GAAVNAFNSNQIVFPAGILQPPFSEKQYQYGNYSWDLADEQNVNGFNTLGTADNG
                NFNGNMMDWNSFSTQHFREQSECMYQYGNYSWDLADEQNVNGFNTLGTADNG
                GVROAYKAYLKWMAEGKQDQLPGLDITHEQLFFINTAQAQWCGSYRPEFAIOSIKTDV
                HSPLKYRVLSLQNLAAAFATFHCACTGPMHPRCVRW"
BASE COUNT 601 a 759 c 813 g 462 t 1 others
ORIGIN

Query Match      100.0%; Score 20; DB 6; Length 2636;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
    |||||
Db 629 CGATGAGGACGCGCTGTTG 610

RESULT 8

```

```
AXI39745/c
LOCUS AXI39745 2663 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 5 from Patent EP1069188.
ACCESSION AXI39745
VERSION AXI39745.1 GI:14275327
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 2663)
Jagerschmidt, A., Agnel, M. and Culouscou, J. M.
Three neprilysin-like membrane metalloproteinases
Patent: EP 1069188-A 5 17-JAN-2001;
SANOFI-SYNTHELABO (FR)
FEATURES
source
1..2663
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..7
8..2296
/feature="unnamed protein product"
/codon_start=1
/translation="MGKSEGVGMVESAGRQKRPGLGGLLLLLLVTAALVALG
VLYADRRGKQLPRLASRLCFLQERTFVKRPRGIPPEAQEVSEVCTTGCVIAARIL
QNMPTPEPCDDFFQACGGWLRHVPETNSRSIFDVLDELEVLKAVLENSTAK
DRPAVERKARTLYRSCMNQSVLEKRGSPQLDLILEVGGWPMVDMRNETVGLWELERQAMNSQF
PLHLERPEQPAHLHPTLQKATVPQERHDVIALYHRMGLFELQSGQLGFDW
LPRDSCVLAQVDELTQKATVPQERHDVIALYHRMGLFELQSGQLGFDW
TLFTQTLSSVKIKLLPDEEVVYGIPLYQNLNIDITYSARTQNYLVRLVLDRIQ
SLSORFKDTRVYRKALGTWVEVRWRECVYVNSNMENAVGSLYVREAPGDSKM
VRELIDKRVTVFETLDELGMDESKKAKQEKAMSIREQIGHDPDIILEENRRIDEE
YSNLFSEDLFENSLQNLKVGAKRSKRLKREKVDNPNLWIIIGAAVNVAFYSPNRRQIV
YPAGILOPPFESKQPNLNGFGIGGMVIGHETITGFDNGRDNKNGMMPWNSFT
QHFRQSECMITQYQNSWDLADQNVNFTLGENIADNGVGRQAYKAYLKWMAEGG
KDQOLPLGLDTHQOLFFINYAQVWCGSYRPEFAIOSIKTDVHSPKRYVLGSLQNLAA
FADTFHCARGTPMHPKRCRVW"
2294..2663
2294..2663
3'UTR 609 a 769 c 816 g 468 t 1 others
BASE COUNT
ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 2663;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CGATGAGGACGCGCTGTG 20
|||||
Db 660 CGATGAGGACGCGCTGTG 641
|||||

RESULT 9
AX033274/c
LOCUS AX033274 2676 bp DNA linear PAT 21-SEP-2000
DEFINITION Sequence 14 from Patent WO0047750.
ACCESSION AX033274
VERSION AX033274.1 GI:10280089
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 2676)
Boileau, G. and Desgroselliers, L.
New metalloproteinases of the neprilysin family
Patent: WO 0047750-A 14 17-AUG-2000;
BOILEAU GUY (CA); DESGROSSELLERS LUC (CA); UNIVERSITE DE MONTREAL
(CA)
FEATURES
source
1..2676
/organism="Homo sapiens"
/db_xref="taxon:9606"
100.0%; Score 20; DB 6; Length 2676;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CGATGAGGACGCGCTGTG 20
|||||
Db 660 CGATGAGGACGCGCTGTG 641
|||||

7..2319
/feature="unnamed protein product"
/codon_start=1
/db_xref="GI:10280090"
/translation="MVESAGRQKRPGLGGLLLLLLVTAALVALGVLVADRRGK
QPLRLASRLCFLQERTFVKRPRGIPPEAQEVSEVCTTGCVIAARIL
CDDYQFACGGWLRHVPETNSRSIFDVLDELEVLKAVLENSTAKDRPAVERKAR
TLYRSCMNQSVLEKRGSPQLDLILEVGGWPMVDMRNETVGLWELERQAMNSQF
NRRVLIDLFINDDQNSRHHIYIDQTLGMPREYIFNGSGNRKRYEALQFMWSVA
TLREDANLPRDSCVLAQVDELTQKATVPQERHDVIALYHRMGLFELQSGQ
FGLGFNWLFTQTLSSVKIKLLPDEEVVYGIPLYQNLNIDITYSARTQNYLVW
RLVLDRIQSLQNRKALGTWVEVRWRECVYVNSNMENAVGSLYVREAPGDSKM
VRELIDKRVTVFETLDELGMDESKKAKQEKAMSIREQIGHDPDIILEENRRIDEE
MNRRLDEEYNSLNFSEDLFENSLQNLKVGAKRSKRLKREKVDNPNLWIIIGAAVNVAFY
SPNRRQIVFPAGILOPPFESKQPNLNGFGIGGMVIGHETITGFDNGRDNKNGM
DWNSFTQSECMITQYQNSWDLADQNVNFTLGENIADNGVGRQAYKAYLKWMAEGG
LKWMAEGGKDQOLPLGLDTHQOLFFINYAQVWCGSYRPEFAIOSIKTDVHSPKRYVL
GSLQNLAAFDTFHCARGTPMHPKRCRVW"
608 a 771 c 823 g 474 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 2676;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CGATGAGGACGCGCTGTG 20
|||||
Db 679 CGATGAGGACGCGCTGTG 660
|||||

RESULT 10
AXI39741/c
LOCUS AXI39741 2714 bp DNA linear PAT 30-MAY-2001
DEFINITION Sequence 1 from Patent EP1069188.
ACCESSION AXI39741
VERSION AXI39741.1 GI:14275323
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 2714)
Jagerschmidt, A., Agnel, M. and Culouscou, J. M.
Three neprilysin-like membrane metalloproteinases
Patent: EP 1069188-A 1 17-JAN-2001;
SANOFI-SYNTHELABO (FR)
FEATURES
source
1..2714
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..7
8..2347
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAC39974.1"
/db_xref="GI:14275324"
/translation="MGKSEGVGMVESAGRQKRPGLGGLLLLLLVTAALVALG
VLYADRRGKQLPRLASRLCFLQERTFVKRPRGIPPEAQEVSEVCTTGCVIAARIL
QNMPTPEPCDDFFQACGGWLRHVPETNSRSIFDVLDELEVLKAVLENSTAK
DRPAVERKARTLYRSCMNQSVLEKRGSPQLDLILEVGGWPMVDMRNETVGLWELER
QALANSPQNRVRLIDLFINDDQNSRHHIYIDQTLGMPREYIFNGSGNRKRYEAL
QFMWSVATLREDANLPRDSCVLAQVDELTQKATVPQERHDVIALYHRMGLFELQSG
MGLEELQSGFLGKGFDTLFTQTLSSVKIKLLPDEEVVYGIPLYQNLNIDITYSA
RTIONLVRLVLDRIQSLQNRKALGTWVEVRWRECVYVNSNMENAVGSLYVREAPGDSKM
VRELIDKRVTVFETLDELGMDESKKAKQEKAMSIREQIGHDPDIILEENRRIDEE
MNRRLDEEYNSLNFSEDLFENSLQNLKVGAKRSKRLKREKVDNPNLWIIIGAAVNVAFY
SPNRRQIVFPAGILOPPFESKQPNLNGFGIGGMVIGHETITGFDNGRDNKNGM
DWNSFTQSECMITQYQNSWDLADQNVNFTLGENIADNGVGRQAYKAYLKWMAEGG
LKWMAEGGKDQOLPLGLDTHQOLFFINYAQVWCGSYRPEFAIOSIKTDVHSPKRYVL
GSLQNLAAFDTFHCARGTPMHPKRCRVW"
621 a 780 c 836 g 476 t 1 others
BASE COUNT
3'UTR 2348..2713
ORIGIN
```

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2714;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCCCTGTTG 20  
 |||||  
 Db 707 CGATGAGGAGCGCCCTGTTG 688

## RESULT 11

AX014701/c  
 LOCUS AX014701  
 DEFINITION Sequence 1 from Patent WO9953077.  
 ACCESSION AX014701  
 VERSION AX014701.1 GI:10040975

## KEYWORDS

SOURCE black rat.  
 ORGANISM Rattus rattus  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

## REFERENCE

1 (bases 1 to 2765)  
 Schwartz,J.C., Gros,C., Ouilmet,T., Rose,C., Bonhomme,M.C. and  
 Facchinetti,P.

## TITLE

Novel nep ii membrane metalloprotease and its use for screening  
 inhibitors-useful in therapy

## JOURNAL

Patent: WO 9953077-A 1 21-OCT-1999;  
 INSTE NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
 CLAUDE (FR); OUILMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
 MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

## FEATURES

source Location/Qualifiers  
 1..2765  
 /organism="Rattus rattus"  
 /db\_xref="taxon:10117"  
 /note="unlabeled protein product"  
 /codon\_start=1  
 /protein\_id="CAC07576.1"  
 /db\_xref="GI:10040975"

## CDS

107..2431  
 /translation="MGKSSVSGMMDERADNCGRRRLGFEVCGLLVLTLLMGAIVTL  
 GFVYSIGKLPILLNLSHVSHERVTVKRVLRDSSOKSDICTTPSCVIAAAILQND  
 QSKKPCDNFQVACGWLHSHVTPETNSRVSFVLDLDELEVILKGVLEDSVQHRPA  
 VKAKTLVASCNMQSVIERKDSPLNLDIMGWPAMDKNETMGPKWELERQLAV  
 LNSQFNRRVLDLFWNDQNSRHVIYIDQPLGMPRSREYFKEDSHRVREAYLQFM  
 TSVATMLRDLNLPGETDLVQEEAAQVLLHLETHANATVQEKRDHVDYALHRMGLAE  
 LQERFLGKFWNTLFTQNVLSVQVVELLPNEVVVYGYPLENLEETIDVFPATLQ  
 YLVRLVLDRIIGLSORFEARVDYRKALYGTTEVVRWRECVSYVNSNMSAVGSLY  
 IKRAFSDKSIIVSELIEKIRSVFVNDLDELNMDDESKKAEKALNIREQIGPYD  
 ILEDNNRHLDSEYSSITFSEDLFFENGLQNKNAQSKLREKVDONLWIIIGAAV  
 NAFYSNRLVFPAGILQPPFPKDPQALNFGIGMVIGHEITHGFDNDGRNFDKN  
 GNMLDWMSNFSARFRQOCMIYQYSNFSELADNQNVNGFTLGENIADNGVROA  
 YKAYLQWLAEGRRDRLGLNLTYYAQLFFINAYQVWCGSYRFEAFQSIKTVDHSPK  
 YRVLSQNLPLGFSFAFHCPGSPMPHMRCLW"

## BASE COUNT

684 a 735 c 787 g 559 t

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCCCTGTTG 20  
 |||||  
 Db 794 CGATGAGGAGCGCCCTGTTG 775

## RESULT 12

AF336981/c  
 LOCUS AF336981  
 DEFINITION Homo sapiens neprilysin-like metalloproteinase 2 mRNA, complete cds.  
 ACCESSION AF336981  
 VERSION AF336981.1 GI:15811370

## KEYWORDS

SOURCE Homo sapiens.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS 1 (bases 1 to 2784)  
 Bonvouloir,N., Lemieux,N., Crine,P., Boileau,G. and  
 DesGroseillers,L.

TITLE Molecular cloning, tissue distribution, and chromosomal  
 localization of MME2, a gene coding for a novel human member of  
 the neutral endopeptidase-24.11 family

JOURNAL DNA Cell Biol. 20 (8), 493-498 (2001)

MEDLINE 21444797  
 PUBMED 11560781

REFERENCE 2 (bases 1 to 2784)  
 Bonvouloir,N., Lemieux,N., Crine,P., Boileau,G. and  
 DesGroseillers,L.

AUTHORS Direct Submission  
 TITLE Submitted (15-JAN-2001) Biochemistry, University of Montreal, P. O.  
 Box 6128, Station Centre Ville, Montreal, Quebec H3C 3J7, Canada

JOURNAL Location/Qualifiers

FEATURES source  
 1..2784

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="1p36"

65..2377  
 /note="NEP-like 2; zinc metalloproteinase"

/codon\_start=1  
 /product="neprilysin-like metalloproteinase 2"

/protein\_id="AAL08942.1"  
 /db\_xref="GI:15811371"

/translation="MVESAGRAGOKRPFLEGLLLVLTLLVLAALVGLVYADRGG  
 QLPRLASRLAFLOEQRVFKRPGIPEAQEVSEVCTPPGCVIAAAILQNDPTEP  
 CDDPYQACGWLHSHVTPETNSRVSFVLDLDELEVILKGVLEDSVQHRPA  
 TLYRSCNMQSVIERKDSPLNLDIMGWPAMDKNETMGPKWELERQLAMNSQF  
 NRVLDLFWNDQNSRHVIYIDQPLGMPRSREYFKEDSHRVREAYLQFM  
 TLLREDANLPRDSCLVEDMVQVLETFQLAKATVPQERHDVIALYHRMGLAEQSLQ  
 FLKAGFNWTLFTQNVLSVQVVELLPNEVVVYGYPLENLEETIDVFPATLQ  
 RLVLDRIGLSORFEARVDYRKALYGTTEVVRWRECVSYVNSNMSAVGSLYREA  
 FPGDSKSVRELIDKRVTFVETDELGMDESKKAEKALNIREQIGPYDILEE  
 MNRDLSEYNLSEDLXFNENGLQNKNAQSKLREKVDPNLWIIIGAAVNAFY  
 DWNSFTQHFRESECMYQYGNISMDLADENQNVNGFTLGENIADNGVROA  
 LKWAEGGKQQLPGLDTHLEFFINAYQVWCGSYRFEAFQSIKTVDHSPKRVLL  
 GSNLQALAAFDTHFCARGTPMHPKRCRW"

BASE COUNT 669 a 789 c 841 g 485 t

ORIGIN

Query Match 100.0%; Score 20; DB 9; Length 2784;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCCCTGTTG 20  
 |||||  
 Db 737 CGATGAGGAGCGCCCTGTTG 718

RESULT 13

AK093058/c

LOCUS AK093058

DEFINITION Homo sapiens cDNA FLJ35739 fls, clone TEST12003824, moderately  
 similar to Mus musculus neprilysin-like metalloproteinase 1 (N11)

ACCESSION AK093058

VERSION AK093058.1 GI:21751802

KEYWORDS oligo capping; fls (full insert sequence).

SOURCE Homo sapiens testis cDNA to mRNA, clone TEST12003824.  
 clone:TEST12003824.

ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1

AUTHORS Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Takahashi-Fuji, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project  
Unpublished  
TITLE NEDO human cDNA sequencing project  
JOURNAL  
REFERENCE 2 (bases 1 to 2850)  
AUTHORS Isogai, T. and Yamamoto, J.  
JOURNAL Direct Submission  
TITLE Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
source 1..2850  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TEST12003824"  
/tissue\_type="testis"  
/clone\_lib="TEST12"  
/note="cloning vector: pME18SFL3"  
BASE COUNT 634 a 847 c 870 g 499 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 9; Length 2850;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCCTGTTG 20  
|||||  
Db 862 CGATGAGGACGCGCCTGTTG 843  
RESULT 14  
AX356951/c 2893 bp DNA linear PAT 13-FEB-2002  
LOCUS  
DEFINITION Sequence 1 from Patent WO0206492.  
ACCESSION AX356951  
VERSION AX356951.1 GI:18674150  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Harrow, I.D., Stacey, P.P., Walsh, R.T., Wayman, C.P., Wayman, C.P. and Phillips, S.C.  
TITLE A novel human soluble secreted endopeptidase (sep) for the treatment of sexual dysfunction  
JOURNAL Patent: WO 0206492-A.1 24-JAN-2002.  
FEATURES  
source 1..2893  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 658 a 854 c 876 g 505 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 2893;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCCTGTTG 20  
|||||  
Db 884 CGATGAGGACGCGCCTGTTG 865  
RESULT 15  
AX463057/c 2893 bp DNA linear PAT 15-JUL-2002  
LOCUS  
DEFINITION Sequence 4 from Patent WO0247670.  
ACCESSION AX463057  
VERSION AX463057.1 GI:21886072  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Naylor, A.M., van der Graaf, P.H. and Wayman, C.P.  
TITLE Treatment of male sexual dysfunction  
JOURNAL Patent: WO 0247670-A.4 20-JUN-2002;  
FEATURES  
source 1..2893  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 658 a 854 c 876 g 505 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 2893;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCCTGTTG 20  
|||||  
Db 884 CGATGAGGACGCGCCTGTTG 865  
Search completed: July 8, 2003, 03:34:49  
Job time : 222.098 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 125.659 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-13

Perfect score: 20

Sequence: 1 cgaatgagcgcgcctgttg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: N\_Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1.	20	100.0	20	AA228819	Rat membrane metal
C 2	20	100.0	2076	AAF89737	Nucleotide sequenc
C 3	20	100.0	2232	AA597186	Human metalloprote
C 4	20	100.0	2262	AAF89739	Nucleotide sequenc
C 5	20	100.0	2286	AAD28130	Soluble secreted e
C 6	20	100.0	2318	AAD30580	Human protease, PR
C 7	20	100.0	2340	AAF89738	Nucleotide sequenc
C 8	20	100.0	2580	ABN84280	Human SEP endopept
C 9	20	100.0	2636	AAF59660	Human neprilysin-1

C 10	20	100.0	2663	22	AAF59661	Human neprilysin-1
C 11	20	100.0	2676	21	AAAG3764	cDNA encoding neut
C 12	20	100.0	2714	22	AAF59659	Human neprilysin-1
C 13	20	100.0	2765	20	AAZ28810	Rat membrane metal
C 14	20	100.0	2893	24	ABN84279	Human SEP endopept
C 15	20	100.0	2893	24	AAD28544	Human soluble secr
C 16	20	100.0	2953	24	ABK48251	cDNA encoding nove
C 17	20	100.0	2975	24	ABK48254	Human SEP cDNA inc
C 18	18.4	92.0	815	23	AA593669	DNA encoding novel
C 19	18.4	92.0	2925	21	AAAG3763	cDNA encoding neut
C 20	16	80.0	1631	21	AAAC33659	Arabidopsis thalia
C 21	15.8	79.0	773	23	ABL05127	Drosophila melanog
C 22	15.8	79.0	1236	21	AAZ53099	Neisseria meningit
C 23	15.8	79.0	1311	21	AAZ53100	Neisseria meningit
C 24	15.8	79.0	1860	24	ABK75489	Bacillus lichenifo
C 25	15.8	79.0	1908	24	ABK75423	Bacillus lichenifo
C 26	15.8	79.0	3083	23	ABL05126	Drosophila melanog
C 27	15.8	79.0	3811	24	AAD28061	Human adenyl and
C 28	15.8	79.0	8307	23	ABL23161	Drosophila melanog
C 29	15.8	79.0	9501	23	AA59550	Propionibacterium
C 30	15.8	79.0	10144	24	ABN95670	Gene #2188 used to
C 31	15.8	79.0	11597	23	ABL23160	Drosophila melanog
C 32	15.8	79.0	47475	21	AAAF1465	N. meningitidis pa
C 33	15.8	79.0	34980	21	AAAF1612	Neisseria meningit
C 34	15.8	79.0	837096	21	AAAB1489	N. meningitidis pa
C 35	15.2	76.0	39	14	AAQ53540	Sequence of the 5'
C 36	15.2	76.0	162	21	AAAB2038	N. meningitidis pa
C 37	15.2	76.0	207	22	ABA50025	Human breast cell
C 38	15.2	76.0	207	22	ABA67950	Human foetal liver
C 39	15.2	76.0	207	22	ABA34994	Probe #13460 for g
C 40	15.2	76.0	207	22	AAK16344	Human brain expres
C 41	15.2	76.0	207	22	AAK42094	Human bone marrow
C 42	15.2	76.0	207	22	AAI22860	Probe #12793 for g
C 43	15.2	76.0	207	22	AAI48159	Probe #16845 used
C 44	15.2	76.0	207	22	AAI08529	Probe #8520 used t
C 45	15.2	76.0	207	24	ABS16136	Human genome deriv

#### ALIGNMENTS

RESULT 1  
AAZ28819  
ID AAZ28819 standard; DNA; 20 BP.  
XX AC AAZ28819;  
XX DT 01-FEB-2000 (first entry)  
XX DE Rat membrane metalloprotease NEPII gene probe #9.  
XX KW Rat; membrane metalloprotease; neprilysin II; NEPII; inactivation; ss;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
XX OS Synthetic.  
XX OS Rattus rattus.  
XX PN FR2777291-A1.  
XX PD 15-OCT-1999.  
XX PF 08-APR-1998; 98FR-0004389.  
XX PR 08-APR-1998; 98FR-0004389.  
XX FA (INRM ) INSERM INST-NAT SANTE & RECH MEDICALE.  
XX PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
XX WPI: 1999-593429/51.

XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease  
XX  
XX Claim 3; Page 22; 29pp; French.  
XX  
XX Sequences AA228811-228827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysin II (NEPII) gene  
CC (AA228810). NEPII is involved in (in)activation of neuronal and hormonal  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
CC disorders.  
XX  
XX Sequence 20 BP; 3 A; 5 C; 8 G; 4 T; 0 other;  
SQ  
Query Match 100.0%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGAGCGCGCTGTG 20  
DB 1 CGATGAGGAGCGCGCTGTG 20  
RESULT 2  
AAF89737/c  
ID AAF89737 standard; DNA; 2076 BP.  
XX  
AC AAF89737;  
XX  
DT 23-JUL-2001 (first entry).  
DE  
DE Nucleotide sequence of a human metalloprotease enzyme IGS5.  
XX  
KW Metalloprotease; IGS5; infection; pain; cancer; diabetes; obesity;  
KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
KW hypotension; hypertension; urinary retention; osteoporosis;  
KW angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
KW benign prostatic hypertrophy; migraine; psychotic disorder;  
KW neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
KW neurodegenerative disease; sleep disorder; epilepsy; kidney disease;  
KW cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
KW subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
KW peripheral vascular disease; Raynaud's disease; motility disorder;  
KW gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
KW inflammation; chemotherapy induced injury; tumour invasion;  
KW immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
KW severe mental retardation; dyskinesia; Huntington's disease;  
KW Gilles de la Tourette's syndrome; ss.  
XX  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2076  
FT FT /\*tag= a  
FT FT /product= "metalloprotease enzyme IGS5"  
XX  
XX WO200136610-A1.  
XX  
XX 25-MAY-2001.  
XX  
XX 17-NOV-2000; 2000WO-EPI1532.  
XX  
XX 19-NOV-1999; 99EP-0203862.  
XX 19-NOV-1999; 99NL-1013616.  
XX 31-MAY-2000; 2000EP-0201937.  
XX 31-MAY-2000; 2000NL-1015356.  
XX

PA (SOLV ) SOLVAY PHARM BV.  
XX  
PI Deleersnijder W, Wiegers R, Weske M;  
XX  
DR WPI; 2001-343815/36.  
DR P-PSDB; AAB83840.  
XX  
XX New IGS5 polypeptides useful for treating infections, pain, cancer,  
PT diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
PT hypertension, urinary retention and Parkinson's disease  
XX  
XX Claim 11; Page 5-6; 115pp; English.  
XX  
XX The present sequence encodes a human metalloprotease enzyme designated  
CC IGS5. IGS5 polynucleotides and polypeptides are useful for treating  
CC infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
CC urinary retention, osteoporosis, angina pectoris, myocardial infarction,  
CC stroke, ulcers, allergies, benign prostatic hypertrophy, migraine,  
CC psychotic and neurological disorders, autism, multiple sclerosis,  
CC Alzheimer's disease, and other neurodegenerative diseases, sleep  
CC disorder, epilepsy, cardiovascular diseases, arteriosclerosis,  
CC cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, kidney  
CC infarction, peripheral vascular disease, Raynaud's disease, kidney  
CC diseases, gastrointestinal disorders, motility disorders and conditions  
CC of delayed gastric emptying, post-operative or diabetic gastroparesis,  
CC diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
CC immune disorders, arthritis, endotoxin shock, sepsis, complications of  
CC diabetes mellitus, and severe mental retardation and dyskinesias, such  
CC as Huntington's disease or Gilles de la Tourette's syndrome.  
XX  
XX Sequence 2076 BP; 493 A; 578 C; 636 G; 369 T; 0 other;  
SQ  
Query Match 100.0%; Score 20; DB 22; Length 2076;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGAGCGCGCTGTG 20  
DB 436 CGATGAGGAGCGCGCTGTG 417  
RESULT 3  
AAS97186/c  
ID AAS97186 standard; cDNA; 2232 BP.  
XX  
AC AAS97186;  
XX  
DT 26-FEB-2002 (first entry)  
DE Human metalloprotease partial DNA sequence #15.  
XX  
KW Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;  
KW vasotropic; antimigraine; analgesic; endocrine; neurotropic; tranquiliser;  
KW hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;  
KW anorectic; antiinflammatory; aspartyl protease; cysteine protease;  
KW metalloprotease; serine protease; cancer; haematopoietic; breast; colon;  
KW lung; prostate; cervical; brain; ovarian; bladder; kidney; pain;  
KW immune-related disease; cardiovascular disease; neuronal disease;  
KW migraine; sexual dysfunction; mood disorder; attention disorder;  
KW cognition disorder; hypotension; hypertension; psychotic disorder;  
KW dyskinesia; metabolic disorder; inflammatory disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200183782-A2.  
XX  
XX 08-NOV-2001.  
XX  
XX 04-MAY-2001; 2001WO-US14431.  
XX  
XX 04-MAY-2000; 2000US-201879P.  
XX

(SUGEN-) SUGEN INC.  
 Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;  
 Payne V;  
 WPI; 2002-041502/05.  
 P-PSDB; AAU72903.  
 Novel protease polypeptide useful for screening for substances that may  
 be used to treat, e.g., cancers, immune-related diseases,  
 cardiovascular disease, migraine, pain, psychotic and inflammatory  
 disorders -  
 Claim 30; Figure 1R-S; 232pp; English.  
 The invention relates to an isolated, enriched, or purified protease  
 polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to  
 screen for substances (S) that may modulate its activity. Administering  
 S (which modulates protease activity in vitro) may be used to treat a  
 disease or disorder selected from cancers (e.g., of tissues, of blood or  
 haematopoietic origin, of the breast, colon, lung, prostate, cervical,  
 brain, ovarian, bladder or kidney), immune-related diseases and  
 disorders, cardiovascular disease, brain or neuronal-associated diseases  
 (e.g., central or peripheral nervous system diseases, migraine, pain,  
 sexual dysfunction, mood disorders, attention disorders, cognition  
 disorders, hypotension, hypertension, psychotic disorders, neurological  
 disorders and dyskinesias), metabolic disorders and inflammatory  
 disorders. (I) may also be useful as a diagnostic tool for a disease or  
 disorder such as those above. AAS97159-AAS97195 represent human  
 protease coding sequences and primers of the invention.  
 Sequence 2232 BP; 512 A; 620 C; 705 G; 395 T; 0 other;  
 Query Match 100.0%; Score 20; DB 24; Length 2232;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGATGAGGACGGCGCTGTG 20  
 |||||  
 Db 595 CGATGAGGACGGCGCTGTG 576  
 RESULT 4  
 AAF89739/c  
 ID AAF89739 standard; DNA; 2262 BP.  
 XX AAF89739;  
 AC AAF89739;  
 XX 23-JUL-2001 (first entry)  
 XX Nucleotide sequence of a human metalloprotease enzyme IG55.  
 XX Metalloprotease; IG55; infection; pain; cancer; diabetes; obesity;  
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
 KW hypotension; hypertension; urinary retention; osteoporosis;  
 KW angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
 KW benign prostatic hypertrophy; migraine; psychotic disorder;  
 KW neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
 KW neurodegenerative disease; sleep disorder; epilepsy; kidney disease;  
 KW cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
 KW subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
 KW peripheral vascular disease; Raynaud's disease; motility disorder;  
 KW gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
 KW inflammation; chemotherapy induced injury; tumour invasion;  
 KW immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
 KW severe mental retardation; dyskinesia; Huntington's disease;  
 KW Gilles de la Tourette's syndrome; ss.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 Key Location/Qualifiers  
 PH 1..2262  
 FT CDS  
 FT /\*tag= a

/product= "metalloprotease enzyme IG55"  
 WO200136610-A1.  
 25-MAY-2001.  
 17-NOV-2000; 2000WO-EP11532.  
 19-NOV-1999; 99EP-0203862.  
 19-NOV-1999; 99NL-1013616.  
 31-MAY-2000; 2000EP-0201937.  
 31-MAY-2000; 2000NL-1015356.  
 (SOLV ) SOLVAY PHARM BV.  
 Deleersnijder W, Wiegers R, Weske M;  
 WPI; 2001-343815/36.  
 P-PSDB; AAB83842.  
 New IG55 polypeptides useful for treating infections, pain, cancer,  
 diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
 hypertension, urinary retention and Parkinson's disease -  
 Claim 11; Page 8-9; 115pp; English.  
 The present sequence encodes a human metalloprotease enzyme designated  
 IG55. IG55 polynucleotides and polypeptides are useful for treating  
 infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 Parkinson's disease, acute heart failure, hypotension, hypertension,  
 urinary retention, osteoporosis, angina pectoris, myocardial infarction,  
 stroke, ulcers, allergies, benign prostatic hypertrophy, migraine,  
 psychotic and neurological disorders, autism, multiple sclerosis,  
 Alzheimer's disease, and other neurodegenerative diseases, sleep  
 disorder, epilepsy, cardiovascular diseases, arteriosclerosis,  
 cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral  
 infarction, peripheral vascular disease, Raynaud's disease, kidney  
 diseases, gastrointestinal disorders, motility disorders and conditions  
 of delayed gastric emptying, post-operative or diabetic gastroparesis,  
 diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
 immune disorders, arthritis, endotoxin shock, sepsis, complications of  
 diabetes mellitus, and severe mental retardation and dyskinesias, such  
 as Huntington's disease or Gilles de la Tourette's syndrome.  
 Sequence 2262 BP; 520 A; 628 C; 716 G; 398 T; 0 other;  
 Query Match 100.0%; Score 20; DB 22; Length 2262;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGATGAGGACGGCGCTGTG 20  
 |||||  
 Db 622 CGATGAGGACGGCGCTGTG 603  
 RESULT 5  
 AAD28130/c  
 ID AAD28130 standard; DNA; 2286 BP.  
 XX AAD28130;  
 AC AAD28130;  
 XX 07-MAY-2002 (first entry)  
 XX Soluble secreted endopeptidase (SEP) consensus DNA.  
 XX Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
 KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
 KW FSD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
 XX Homo sapiens.  
 OS Mus sp.  
 OS Rattus sp.

XX FH Key Location/Qualifiers  
FT misc\_feature 1664..2286  
FT /tag= a  
FT /note= "Encodes catalytic domain".  
XX PN W0200206492-A1.  
XX PD 24-JAN-2002.  
XX 16-JUL-2001; 2001WO-IB01263.  
XX 14-JUL-2000; 2000GB-0017387.  
XX (PFIZ ) PFIZER LTD.  
XX (PFIZ ) PFIZER INC.  
XX Harrow ID., Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
XX WPI; 2002-155042/20.  
XX An isolated and/or purified nucleic acid encoding a human soluble  
PT secreted endopeptidase which is useful for treating sexual dysfunction,  
PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
PT female sexual arousal disorder -  
XX Disclosure; Fig 6; 167pp; English.  
XX The invention relates to an isolated and/or purified nucleic acid  
CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
CC antibody and the compound which inhibits or selectively inhibits the  
CC human SEP protein are useful in the manufacture of a medicament for  
CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
CC also useful for treating the above disorders and other disorders such  
CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
CC hypoactive sexual desire disorder. The present sequence is SEP consensus  
CC DNA sequence found in human, mouse and rat.  
XX SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
Query Match 100.0%; Score 20; DB 24; Length 2286;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGACGCGCTGTG 20  
|||||  
DB 670 CGATGAGGACGCGCTGTG 651  
RESULT 6  
AAD30580/c  
ID AAD30580 standard; cDNA: 2318 BP.  
XX AC AAD30580;  
XX 21-MAY-2002 (first entry)  
XX Human protease, PRS-13 CDNA.  
XX Human; protease; PRS-13; enzyme; gastritis; cirrhosis; Crohn's disease;  
KW gastrointestinal disorder; autoimmune; inflammatory; cell proliferative;  
KW cardiovascular; developmental; epithelial; neurological; reproductive;  
KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;  
KW anaemia; asthma; atherosclerosis; hypertension; myocardial infarction;  
KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;  
KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;  
KW Pick's disease; infertility; vitiligo; drug screening; gene therapy; ss.  
XX Homo sapiens.  
XX OS  
XX FH Key Location/Qualifiers

FT CDS 338..1651  
FT /tag= a  
FT /product= "Human PRS-13 protein"  
FT sig\_peptide 338..427  
FT /tag= b  
FT mat\_peptide 428..1648  
FT /tag= c  
FT /product= "Mature PRS-13 protein"  
XX WO200208396-A2.  
XX 31-JAN-2002.  
XX 17-JUL-2001; 2001WO-US22397.  
XX 21-JUL-2000; 2000US-220063P.  
XX 28-JUL-2000; 2000US-221680P.  
XX 04-AUG-2000; 2000US-223544P.  
XX 11-AUG-2000; 2000US-224717P.  
XX 16-AUG-2000; 2000US-225988P.  
XX 23-AUG-2000; 2000US-227568P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Delegeane AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;  
PI Tribouley CM, Das D, Kallifin JA, Policky JL, Ramkumar J, Yang J;  
PI Yue H, Au-Young J, Griffin JA, Kearney L, Baughn MR, Borowsky ML;  
PI Thangavelu K, Ding L, Kearney L, Burford N, Walia NK, Lal P, Lee S, Todd S;  
PI Sanjanwala MS, Yao MG, Burford N, Walia NK, Lal P, Lee S, Todd S;  
PI Lo TP, Tang YT, Elliott VS, Azimzal Y, Lu Y;  
XX WPI; 2002-206082/26.  
XX P-FSDB; AAE19176.  
XX New human protease polypeptide, useful in diagnosis, prevention and  
PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,  
PT cell proliferative, developmental, epithelial and neurological  
PT disorders  
XX Claim 5; Page 174-175; 182pp; English.  
XX The invention relates to an isolated human protease polypeptide (PRS).  
CC PRS protein and DNA are useful for diagnosing, treating and preventing  
CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),  
CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,  
CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,  
CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,  
CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),  
CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders  
CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,  
CC Parkinson's disease), and reproductive disorders (infertility). PRS  
CC protein is useful in a number of drug screening techniques and to  
CC analyse the proteome of a tissue or cell type. PRS DNA is useful for  
CC creating knockin humanised animals or transgenic animals to model human  
CC diseases, in somatic or germline gene therapy and in microarrays  
CC utilising fluids or tissues from patients to detect altered PKIN  
CC expression. The present sequence is human PRS-13 CDNA.  
XX SQ Sequence 2318 BP; 526 A; 647 C; 735 G; 410 T; 0 other;  
Query Match 100.0%; Score 20; DB 24; Length 2318;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGACGCGCTGTG 20  
|||||  
DB 684 CGATGAGGACGCGCTGTG 665  
RESULT 7  
AAF89738/c  
ID AAF89738 standard; DNA: 2340 BP.  
XX

AAF89738;  
23-JUL-2001 (first entry)  
Nucleotide sequence of a human metalloprotease enzyme IGS5.  
Metalloprotease; IGS5; infection; pain; cancer; diabetes; obesity;  
anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
hypertension; hypertension; urinary retention; osteoporosis;  
angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
benign prostatic hypertrophy; migraine; psychotic disorder;  
neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
neurodegenerative disease; sleep disorder; epilepsy; kidney disease;  
cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
peripheral vascular disease; Raynaud's disease; motility disorder;  
gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
inflammation; chemotherapy induced injury; tumour invasion;  
immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
severe mental retardation; dyskinesia; Huntington's disease;  
Gilles de la Tourette's syndrome; ss.  
Homo sapiens.  
Key Location/Qualifiers  
CDS 1..2340  
/tag= a  
/product= "metalloprotease enzyme IGS5"  
WO200136610-A1.  
25-MAY-2001.  
17-NOV-2000; 2000WO-EP11532.  
19-NOV-1999; 99EP-0203862.  
19-NOV-1999; 99NL-1013616.  
31-MAY-2000; 2000EP-0201937.  
31-MAY-2000; 2000NL-1015356.  
(SOLV ) SOLVAY PHARM BV.  
Deleersnijder W, Wieggers R, Weske M;  
WPI; 2001-343815/36;  
P-PSDB; AAB83841.  
New IGS5 polypeptides useful for treating infections, pain, cancer,  
diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
hypertension, urinary retention and Parkinson's disease  
Claim 11; Page 6-7; 115pp; English.  
The present sequence encodes a human metalloprotease enzyme designated  
IGS5. IGS5 polynucleotides and polypeptides are useful for treating  
infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
Parkinson's disease, acute heart failure, hypotension, hypertension,  
urinary retention, osteoporosis, angina pectoris, myocardial infarction,  
stroke, ulcers, allergies, benign prostatic hypertrophy, migraine,  
psychotic and neurological disorders, autism, multiple sclerosis,  
Alzheimer's disease, and other neurodegenerative diseases, sleep  
disorder, epilepsy, cardiovascular diseases, arteriosclerosis,  
cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral  
infarction, peripheral vascular disease, Raynaud's disease, kidney  
diseases, gastrointestinal disorders, motility disorders and conditions  
of delayed gastric emptying, post-operative or diabetic gastroparesis,  
diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
immune disorders, arthritis, endotoxin shock, sepsis, complications of  
diabetes mellitus, and severe mental retardation and dyskinesias, such  
as Huntington's disease or Gilles de la Tourette's syndrome.  
Sequence 2340 BP; 539 A; 649 C; 739 G; 413 T; 0 other;

Query Match 100.0%; Score 20; DB 22; Length 2340;  
Best Local Similarity 100.0%; Pred. No. 2.11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CGATGAGGACGCGCTGTTG 20  
Db 700 CGATGAGGACGCGCTGTTG 681  
|||||  
RESULT 8  
ABN84280/C  
ID ABN84280 standard; cDNA; 2580 BP.  
XX AC ABN84280;  
XX DT 23-SEP-2002 (first entry)  
XX Human SEP endopeptidase coding sequence.  
XX SEP; endopeptidase; enzyme; human; neuropeptide Y1; receptor;  
XX male sexual dysfunction; male erectile dysfunction; obesity;  
XX anorexia; bulimia; vasotropic; anorectic; therapy; inhibitor; gene;  
XX ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH misc\_feature 1..65  
FT /tag= a  
FT /note= "5' partial vector sequence"  
FT CDS 258..2580  
FT /tag= b  
FT /product= "SEP"  
FT /partial  
FT /note= "the CDS does not include a stop codon"  
FT misc\_feature 2580  
FT /tag= c  
FT /note= "the 3' end of the sequence is missing  
from Figure 8 of the specification"  
XX WO200247670-A1.  
XX PN  
XX PD 20-JUN-2002.  
XX 10-DEC-2001; 2001WO-IB02399.  
XX 15-DEC-2000; 2000GB-0030647.  
XX 06-APR-2001; 2001GB-0008730.  
XX 23-APR-2001; 2001GB-0009910.  
XX 04-MAY-2001; 2001GB-0011037.  
XX 29-JUN-2001; 2001US-0895367.  
XX 13-JUL-2001; 2001US-0905846.  
XX 24-AUG-2001; 2001GB-0020679.  
XX (PFIZ ) PFIZER LTD.  
XX (PFIZ ) PFIZER INC.  
XX Naylor AM, Van Der Graaf PH, Wayman CP;  
WPI; 2002-547828/58.  
P-PSDB; ABB79521.  
Use of an inhibitor of neuropeptide Y in the preparation of medicament  
for the treatment or prevention of male erectile dysfunction  
Disclosure; Fig 8; 179pp; English.  
The present sequence is a nucleotide sequence coding for human  
SEP, a soluble secreted endopeptidase, with an additional 5'  
vector partial sequence. The invention relates to the use of an  
inhibitor of NPY, or an inhibitor of NPY Y1 receptor, especially an  
inhibitor selective for an NPY or NPY Y1 receptor associated with  
male genitalia, in the preparation of a medicament for the

CC treatment or prevention of male sexual dysfunction, especially male  
 CC erectile dysfunction (MED). The NPY inhibitor may be used with an  
 CC auxiliary active agent such as an SEP inhibitor. The invention  
 CC provides a SEP assay that can be used to detect candidate  
 CC inhibitors of SEP. In addition to treatment of MED, NPY inhibitors  
 CC can also be used to treat abnormal drink and food intake disorders,  
 CC such as obesity, bulimia, anorexia and metabolic disorders.

XX Sequence 2580 BP; 587 A; 741 C; 794 G; 458 T; 0 other;

SQ Query Match 100.0%; Score 20; DB 24; Length 2580;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGGCCCTGTTG 20

|||||  
 DB 949 CGATGAGGACGGCCCTGTTG 930

RESULT 9

AAF59660/c

ID AAF59660 standard; cDNA; 2636 BP.

XX

AC AAF59660;

XX

DT 27-APR-2001 (first entry)

XX

DE Human neprilysin-like membrane metalloproteinase SNEPB cDNA.

XX

KW Human; SNEPB; neprilysin-like membrane metalloproteinase;

KW splice variant; alternative splicing; zinc endopeptidase family;

KW neurotensin; peptide hormone; processing; metabolism; vaccine;

KW drug screening; renal insufficiency; hepatic disease; pain;

KW hypertension; cancer; inflammation; cardiovascular disease;

KW neuronal disease; pancreatic disease; prostatic disease;

KW respiratory disease; gene therapy; nephrotropic; vasotropic; analgesic;

KW cerebroprotective; hypotensive; cytostatic; antiinflammatory;

KW cardiovascular; hepatotropic; ss.

XX

OS Homo sapiens.

XX

PN EP1069188-A1.

XX

PD 17-JAN-2001.

XX

PF 15-JUL-1999; 99EP-0401767.

XX

PR 15-JUL-1999; 99EP-0401767.

XX

PA (SNFI ) SANOFI-SYNTHELABO.

XX

PI Jagerschmidt A, Agnel M, Culouscou J;

XX

DR WPI; 2001-212582/22.

XX

DR P-PSDB; AAB60562.

XX

PT New membrane-associated metalloproteinase SNEPB, SNEPB and SNEPC

PT polypeptides and polynucleotides, useful for treating e.g. acute and

PT chronic renal insufficiency, pain, stroke, cancer, inflammation, renal

PT and hepatic ischemia

XX

PS Claim 5; Page 30-33; 72pp; English.

XX

CC The invention relates to the human SNEPB, SNEPB and SNEPC proteins, and

CC the cDNAs encoding them. SNEPB, SNEPB and SNEPC are neprilysin-like

CC membrane metalloproteinases and are the products of alternative splicing.

CC The substrate(s) for the SNEPB proteins are not as yet known, although

CC the neprilysin family of zinc endopeptidases play key roles in the

CC processing and/or metabolism of neurotensin and peptide hormones. SNEPB

CC nucleotides may be used as hybridisation probes for cDNA and genomic

CC DNA; to isolate full-length cDNAs and genomic clones encoding SNEPB,

CC SNEPB or SNEPC; to isolate cDNA and genomic clones of SNEPB homologues;

CC as research reagents and material for the discovery of treatments and

CC

CC diagnostics for animal and human diseases; and for chromosome  
 CC identification. The SNEPB proteins may be used as immunogens to  
 CC produce antibodies immunospecific for SNEPB, SNEPB or SNEPC. Such  
 CC antibodies are used to isolate or identify clones expressing the  
 CC protein, or to purify the proteins by affinity chromatography.  
 CC SNEPB proteins may also be used in screening for compounds which modulate  
 CC SNEPB endopeptidase activity, and to assess enzymatic cleavage of small  
 CC molecule substrates in cells, cell-free preparations, chemical libraries  
 CC and product mixtures. The SNEPB proteins (as vaccine compositions),  
 CC SNEPB nucleotides, and SNEPB activators or inhibitors may be used  
 CC to treat acute and chronic renal insufficiency, cancer, inflammation, as  
 CC ischemia, pain, stroke, hypertensive disease, cancer, inflammation, as  
 CC well as cardiovascular, neuronal, pancreatic, prostatic, renal,  
 CC respiratory or hepatic diseases. They may also be used in modulating  
 CC peptide activation and/or degradation in the brain or kidney or in  
 CC another organ, or to diagnose or treat any disorder related to abnormal  
 CC expression of SNEPB, SNEPB or SNEPC. The present sequence represents  
 CC cDNA encoding SNEPB.

XX Sequence 2636 BP; 601 A; 759 C; 813 G; 462 T; 1 other;

Query Match 100.0%; Score 20; DB 22; Length 2636;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGGCCCTGTTG 20

|||||

DB 629 CGATGAGGACGGCCCTGTTG 610

RESULT 10

AAF59661/c

ID AAF59661 standard; cDNA; 2663 BP.

XX

AC AAF59661;

XX

DT 27-APR-2001 (first entry)

XX

DE Human neprilysin-like membrane metalloproteinase SNEPB cDNA.

XX

KW Human; SNEPB; neprilysin-like membrane metalloproteinase;

KW splice variant; alternative splicing; zinc endopeptidase family;

KW neurotensin; peptide hormone; processing; metabolism; vaccine;

KW drug screening; renal insufficiency; hepatic disease; pain;

KW hypertension; cancer; inflammation; cardiovascular disease;

KW neuronal disease; pancreatic disease; prostatic disease;

KW respiratory disease; gene therapy; nephrotropic; vasotropic; analgesic;

KW cerebroprotective; hypotensive; cytostatic; antiinflammatory;

KW cardiovascular; hepatotropic; ss.

XX

OS Homo sapiens.

XX

PN EP1069188-A1.

XX

PD 17-JAN-2001.

XX

PF 15-JUL-1999; 99EP-0401767.

XX

PR 15-JUL-1999; 99EP-0401767.

XX

PA (SNFI ) SANOFI-SYNTHELABO.

XX

PI Jagerschmidt A, Agnel M, Culouscou J;

XX

DR WPI; 2001-212582/22.

XX

DR P-PSDB; AAB60563.

XX

PT New membrane-associated metalloproteinase SNEPB, SNEPB and SNEPC

PT polypeptides and polynucleotides, useful for treating e.g. acute and

PT chronic renal insufficiency, pain, stroke, cancer, inflammation, renal

PT and hepatic ischemia

XX

PS Claim 5; Page 35-38; 72pp; English.

XX

XX CC The invention relates to the human SNEPa, SNEPb and SNEPc proteins, and  
 CC the cDNAs encoding them. SNEPa, SNEPb and SNEPc are neprilysin-like  
 CC membrane metalloproteinases and are the products of alternative splicing.  
 CC The substrate(s) for the SNEP proteins are not as yet known, although  
 CC the neprilysin family of zinc endopeptidases play key roles in the  
 CC processing and/or metabolism of neuropeptides and peptide hormones. SNEP  
 CC nucleotides may be used as hybridisation probes for cDNA and genomic  
 CC DNA; to isolate full-length cDNAs and genomic clones encoding SNEPa,  
 CC SNEPb or SNEPc; to isolate cDNA and genomic clones of SNEP homologues;  
 CC as research reagents and material for the discovery of treatments and  
 CC diagnostics for animal and human diseases; and for chromosome  
 CC identification. The SNEP proteins may be used as immunogens to  
 CC produce antibodies immunospecific for SNEPa, SNEPb or SNEPc. Such  
 CC antibodies are used to isolate or identify clones expressing the  
 CC protein, or to purify the proteins by affinity chromatography.  
 CC SNEP proteins may also be used in screening for compounds which modulate  
 CC SNEP endopeptidase activity, and to assess enzymatic cleavage of small  
 CC molecule substrates in cells, cell-free preparations, chemical libraries  
 CC and product mixtures. The SNEP proteins (as vaccine compositions),  
 CC SNEP nucleotides, and SNEP activators or inhibitors may be used  
 CC to treat acute and chronic renal insufficiency, renal and hepatic  
 CC ischaemia, pain, stroke, hypertensive disease, cancer, inflammation, as  
 CC well as cardiovascular, neuronal, pancreatic, prostatic, renal,  
 CC respiratory or hepatic diseases. They may also be used in modulating  
 CC peptide activation and/or degradation in the brain or kidney or in  
 CC another organ, or to diagnose or treat any disorder related to abnormal  
 CC expression of SNEPa, SNEPb or SNEPc. The present sequence represents  
 CC cDNA encoding SNEPc.

XX SQ Sequence 2663 BP; 609 A; 769 C; 816 G; 468 T; 1 other;

Query Match 100.0%; Score 20; DB:22; Length 2663;  
 Best Local Similarity 100.0%; Pred. NO. 2.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTG 20  
 |||||  
 Db 660 CGATGAGGAGCGCGCTGTG 641

RESULT 11  
 AAA63764/c  
 ID AAA63764 standard; cDNA; 2676 BP.  
 XX AC AAA63764;

XX DT 04-DEC-2000 (first entry)

XX DE cDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-2.

XX KW Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
 KW NEP-like enzyme; protein production; protein secretion;  
 KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
 XX fertility; bone disease; abnormal phosphate metabolism; ss.

OS Homo sapiens.

XX Key Location/Qualifiers  
 FH CDS 7..2319  
 FT /\*tag= a  
 FT /product= "neutral endopeptidase metalloproteinase-like  
 FT enzyme NL-2"

XX PN W0200047750-A2.

XX PD 17-AUG-2000.

XX PE 11-FEB-2000; 2000WO-CA00147.

XX PR 11-FEB-1999; 99CA-2260376.

XX PA (UYMO-) UNIV MONTREAL.

XX PI Desgroseillers L, Boileau G;  
 XX DR WPI; 2000-549148/50;  
 DR P-PSDB; AAB08131.

XX PT Novel neutral endopeptidase-like metalloproteinase polypeptides and  
 PT polynucleotides, used to screen for related sequences and enzyme  
 PT inhibitors, used for the treatment of NL-3 related bone disorders -  
 XX Disclosure; Fig 4; 59pp; English.

XX CC The present sequence encodes a human neutral endopeptidase  
 CC metalloproteinase-like enzyme, designated NL-2. The specification  
 CC also describes NL-1 and NL-3. The NL enzymes are used to test for  
 CC specific inhibitors. The N-terminal region of the enzymes can be used  
 CC to promote production and secretion of foreign proteins and active  
 CC bioproteins, using chimeric constructs containing the foreign protein  
 CC downstream from and in phase with the N-terminal region. The NL enzymes  
 CC are have been localised to the brain, and may be useful in the  
 CC treatment of neurological diseases such as Alzheimer's disease, pain,  
 CC and psychiatric disorders. NL enzymes have also been localised to the  
 CC testis and ovaries, and may be used to control fertility. They have  
 CC also been localised to bones, and may be used to treat bone diseases,  
 CC and abnormal phosphate metabolisms related to improper peptide  
 CC processing by the NL-3 enzyme.

XX SQ Sequence 2676 BP; 608 A; 771 C; 823 G; 474 T; 0 other;

Query Match 100.0%; Score 20; DB 21; Length 2676;  
 Best Local Similarity 100.0%; Pred. NO. 2.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTG 20  
 |||||  
 Db 679 CGATGAGGAGCGCGCTGTG 660

RESULT 12  
 AAF59659/c  
 ID AAF59659 standard; cDNA; 2714 BP.

XX AC AAF59659;

XX DT 27-APR-2001 (first entry)

XX DE Human neprilysin-like membrane metalloproteinase SNEPa cDNA.

XX KW Human; SNEPa; neprilysin-like membrane metalloproteinase;  
 KW splice variant; alternative splicing; zinc endopeptidase family;  
 KW neurotrophic; peptide hormone; processing; metabolism; vaccine;  
 KW drug screening; renal insufficiency; hepatic disease; ischaemia; pain;  
 KW hypertension; cancer; inflammation; cardiovascular disease;  
 KW neuronal disease; pancreatic disease; prostatic disease;  
 KW respiratory disease; gene therapy; nephrotropic; analgesic;  
 KW cerebroprotective; hypotensive; cytostatic; antiinflammatory;  
 KW cardiovascular; hepatotropic; ss.

XX OS Homo sapiens.

XX PN EP1069188-A1.

XX PD 17-JAN-2001.

XX PF 15-JUL-1999; 99EP-0401767.

XX PR 15-JUL-1999; 99EP-0401767.

XX PA (SNFI) SANOFI-SYNTHELABO.

XX PI Jagerschmidt A, Agnel M, Culouscou J;

XX WPI; 2001-212582/22.

DR P-PSDB; AAB60561.

XX New membrane-associated metalloproteinase SNEPa, SNEPb and SNEPc  
PT polypeptides and polynucleotides, useful for treating e.g. acute and  
PT chronic renal insufficiency, pain, stroke, cancer, inflammation, renal  
PT and hepatic ischemia

XX Claim 5; Page 25-28; 72pp; English.

XX The invention relates to the human SNEPa, SNEPb, and SNEPc proteins, and  
XX the cDNAs encoding them. SNEPa, SNEPb and SNEPc are neprilysin-like  
XX membrane metalloproteinases and are the products of alternative splicing.  
XX The substrate(s) for the SNEP proteins are not as yet known, although  
XX the neprilysin family of zinc endopeptidases play key roles in the  
XX processing and/or metabolism of neuro-peptides and peptide hormones. SNEP  
XX nucleotides may be used as hybridisation probes for cDNA and genomic  
XX DNA; to isolate full-length cDNAs and genomic clones encoding SNEPa,  
XX SNEPb or SNEPc; to isolate cDNA and genomic clones of SNEP homologues;  
XX as research reagents and material for the discovery of treatments and  
XX diagnostics for animal and human diseases; and for chromosome  
XX identification. The SNEP proteins may be used as immunogens to  
XX produce antibodies immunospecific for SNEPa, SNEPb or SNEPc. Such  
XX antibodies are used to isolate or identify clones expressing the  
XX protein, or to purify the proteins by affinity chromatography.  
XX SNEP proteins may also be used in screening for compounds which modulate  
XX SNEP endopeptidase activity, and to assess enzymatic cleavage of small  
XX molecule substrates in cells, cell-free preparations, chemical libraries  
XX and product mixtures. The SNEP proteins (as vaccine compositions),  
XX SNEP nucleotides, and SNEP activators or inhibitors may be used  
XX to treat acute and chronic renal insufficiency, renal and hepatic  
XX ischaemia, pain, stroke, hypertensive disease, cancer, inflammation, as  
XX well as cardiovascular, neuronal, pancreatic, prostatic, renal,  
XX respiratory or hepatic diseases. They may also be used in modulating  
XX peptide activation and/or degradation in the brain or kidney or in  
XX another organ, or to diagnose or treat any disorder related to abnormal  
XX expression of SNEPa, SNEPb or SNEPc. The present sequence represents  
XX cDNA encoding SNEPa.

XX Sequence 2714 BP; 621 A; 780 C; 836 G; 476 T; 1 other;

Query Match 100.0%; Score 20; DB 22; Length 2714;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGATGAGGAGCGCCCTGTG 20  
Db 707 CGATGAGGAGCGCCCTGTG 688

RESULT 13

AAZ28810/C  
ID AAZ28810 standard; cDNA; 2765 BP.

XX AAZ28810;

XX 01-FEB-2000 (first entry)

DE Rat membrane metalloproteinase NEPII gene.

XX Rat; membrane metalloproteinase; neprilysin II; NEPII; inactivation;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; ds.

XX Rattus rattus.

XX FR2777291-A1.

XX 15-OCT-1999.

XX 08-APR-1998; 98FR-0004389.

XX 08-APR-1998; 98FR-0004389.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;

XX WPI; 1999-593429/51.

XX P-PSDB; AAY44177.

XX New membrane metalloproteinase NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease

XX Claim 2; Page 12-16; 29pp; French.

XX This sequence represents the gene for the rat membrane metalloproteinase  
CC designated neprilysin II (NEPII), which is involved in (in)activation  
CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
CC for specific substrates (used to detect NEPII in cells and tissues) or  
CC inhibitors, which can also be used to detect NEPII or for treatment of  
CC disorders related to peptidergic signalling in which NEPII is involved,  
CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
CC endocrine disorders.

XX Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;

Query Match 100.0%; Score 20; DB 20; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CGATGAGGAGCGCCCTGTG 20  
Db 794 CGATGAGGAGCGCCCTGTG 775

RESULT 14

ABN84279/C

ID ABN84279 standard; cDNA; 2893 BP.

XX ABN84279;

XX 23-SEP-2002 (first entry)

DE Human SEP endopeptidase coding sequence.

XX SEP; endopeptidase; enzyme; human; neuropeptide Y1; receptor;  
KW male sexual dysfunction; male erectile dysfunction; obesity;  
KW anorexia; bulimia; vasotropic; anorectic; therapy; inhibitor; gene;  
KW ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
CDS 185..2524  
FT /\*tag= a  
FT /product= "SEP"

XX WO200247670-A1.

XX 20-JUN-2002.

XX 10-DEC-2001; 2001WO-IB02399.

XX 15-DEC-2000; 2000GB-0030647.

XX 06-APR-2001; 2001GB-0008730.

XX 23-APR-2001; 2001GB-0009910.

XX 04-MAY-2001; 2001GB-0011037.

XX 19-JUN-2001; 2001US-0895367.

XX 13-JUL-2001; 2001US-0905846.

XX 24-AUG-2001; 2001GB-0020679.

XX (PFIZ ) PFIZER LTD.

PA (PFIZ ) PFIZER INC.  
 XX Naylor AM, Van Der Graaf PH, Wayman CP;  
 PI WPI; 2002-547828/58.  
 XX P-PSDB; ABB79521.  
 DR use of an inhibitor of neuropeptide Y in the preparation of medicament  
 XX for the treatment or prevention of male erectile dysfunction  
 PT Disclosure; Fig 7; 179pp; English.  
 XX  
 PS The present sequence is a nucleotide sequence coding for human  
 XX SEP, a soluble secreted endopeptidase. The invention relates to  
 CC the use of an inhibitor of NPY, or an inhibitor of NPY Y1 receptor,  
 CC especially an inhibitor selective for an NPY or NPY Y1 receptor  
 CC associated with male genitalia, in the preparation of a medicament  
 CC for the treatment or prevention of male sexual dysfunction,  
 CC especially male erectile dysfunction (MED). The NPY inhibitor  
 CC may be used with an auxiliary active agent such as an SEP  
 CC inhibitor. The invention provides a SEP assay that can be used to  
 CC detect candidate inhibitors of SEP. In addition to treatment of  
 CC MED, NPY inhibitors can also be used to treat abnormal drink and  
 CC food intake disorders, such as obesity, bulimia, anorexia and  
 CC metabolic disorders.  
 XX  
 SQ Sequence 2893 BP; 658 A; 854 C; 876 G; 505 T; 0 other;  
 Query Match 100.0%; Score 20; DB 24; Length 2893;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGATGAGGACGCCCTGTTG 20  
 DB 884 CGATGAGGACGCCCTGTTG 865  
 RESULT 15  
 AAD28544/C  
 ID AAD28544 standard; cDNA; 2893 BP.  
 XX  
 AC AAD28544;  
 XX  
 DT 07-MAY-2002 (first entry)  
 XX  
 DE Human soluble secreted endopeptidase (SEP) cDNA.  
 XX  
 KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
 KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
 KW FSAD; premature ejaculation; anorgasmia; vaginismus; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 185..2524  
 FT /\*tag= a  
 FT /product= "Human SEP protein"  
 FT misc\_feature 1711..2893  
 FT /\*tag= b  
 FT /note= "Encodes catalytic domain"  
 XX  
 PN WO200206492-A1.  
 XX  
 PD 24-JAN-2002.  
 XX  
 XX 16-JUL-2001; 2001WO-IB01263.  
 XX  
 PR 14-JUL-2000; 2000GB-0017387.  
 XX  
 XX (PFIZ ) PFIZER LTD.  
 PA (PFIZ ) PFIZER INC.  
 XX

PI Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
 XX WPI; 2002-155042/20.  
 DR P-PSDB; AAE17779.  
 XX  
 PT An isolated and/or purified nucleic acid encoding a human soluble  
 PT secreted endopeptidase which is useful for treating sexual dysfunction,  
 PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
 PT female sexual arousal disorder -  
 XX Claim 1; Fig 1; 167pp; English.  
 XX  
 PS The invention relates to an isolated and/or purified nucleic acid  
 CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
 CC antibody and the compound which inhibits or selectively inhibits the  
 CC human SEP protein are useful in the manufacture of a medicament for  
 CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
 CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
 CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
 CC also useful for treating the above disorders and other disorders such  
 CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
 CC hypoaffective sexual desire disorder. The present sequence is human SEP  
 CC cDNA.  
 XX  
 SQ Sequence 2893 BP; 658 A; 854 C; 876 G; 505 T; 0 other;  
 Query Match 100.0%; Score 20; DB 24; Length 2893;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CGATGAGGACGCCCTGTTG 20  
 DB 884 CGATGAGGACGCCCTGTTG 865  
 Search completed: July 8, 2003, 02:19:00  
 Job time : 126.659 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July. 8, 2003, 01:24:03 : Search time 27.6098 Seconds

(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-13

Perfect score: 20

Sequence: 1 cgaagagacgcctgttg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.2	76.0	39	1	US-08-225-989-9
C 2	15.2	76.0	39	1	US-08-570-923-9
C 3	15.2	76.0	39	1	US-08-580-014-9
C 4	15.2	76.0	39	3	US-09-079-785-9
C 5	15.2	76.0	1434	4	US-09-587-700-1
C 6	15.2	76.0	5633	4	US-09-221-017B-545
C 7	15.2	76.0	31328	4	US-09-215-694-19
C 8	15.2	76.0	4403765	4	US-09-103-840A-2
C 9	15.2	76.0	4411529	4	US-09-103-840A-1
C 10	14.8	74.0	5134	2	US-08-635-121-1
C 11	14.8	74.0	30001	1	US-08-125-468-1
C 12	14.8	74.0	30001	1	US-08-474-933-1
C 13	14.4	72.0	1302	2	US-08-529-600D-1
C 14	14.4	72.0	1302	2	US-08-973-275-5
C 15	14.4	72.0	1302	3	US-09-122-632-1
C 16	14.2	71.0	30	4	US-09-240-179-11
C 17	14.2	71.0	30	4	US-09-240-179-12
C 18	14.2	71.0	423	1	US-08-470-179-98
C 19	14.2	71.0	555	4	US-09-387-800-1
C 20	14.2	71.0	630	4	US-09-199-637A-400
C 21	14.2	71.0	783	4	US-09-134-001C-1670
C 22	14.2	71.0	843	3	US-08-888-429A-25
C 23	14.2	71.0	1191	4	US-09-282-305-13
C 24	14.2	71.0	1215	1	US-08-420-235B-18
C 25	14.2	71.0	1215	3	US-08-793-624-18
C 26	14.2	71.0	1215	5	PCR-US95-10194-18
C 27	14.2	71.0	1283	4	US-09-282-305-11

C 28 14.2 71.0 1308 4 US-09-446-754-1  
C 29 14.2 71.0 1410 2 US-08-343-101A-6  
C 30 14.2 71.0 1410 3 US-09-183-688-6  
C 31 14.2 71.0 1461 3 US-09-344-001-1  
C 32 14.2 71.0 1689 4 US-09-311-924-1  
C 33 14.2 71.0 1929 4 US-09-380-420C-1  
C 34 14.2 71.0 1973 4 US-09-311-924-3  
C 35 14.2 71.0 2094 3 US-09-019-385-1  
C 36 14.2 71.0 2473 4 US-09-173-914-3  
C 37 14.2 71.0 2835 4 US-09-134-001C-1515  
C 38 14.2 71.0 3293 1 US-08-030-096-1  
C 39 14.2 71.0 3561 4 US-09-134-001C-1685  
C 40 14.2 71.0 3597 4 US-09-199-637A-404  
C 41 14.2 71.0 6414 4 US-09-134-001C-1626  
C 42 14.2 71.0 8878 1 US-08-759-444-2  
C 43 14.2 71.0 9880 3 US-08-680-897-1  
C 44 14.2 71.0 20710 1 US-08-420-235B-1  
C 45 14.2 71.0 20710 3 US-08-793-624-1

## ALIGNMENTS

RESULT 1  
US-08-225-989-9/c  
: Sequence 9, Application US/08225989  
: Patent No. 5480981  
: GENERAL INFORMATION:  
: APPLICANT: Goodwin, Raymond G.  
: APPLICANT: Smith, Craig A.  
: APPLICANT: Armitage, Richard J.  
: APPLICANT: Gruss, Hans-Jurgen  
: TITLE OF INVENTION: No. 5480981el Cytokine That Binds CD30  
: NUMBER OF SEQUENCES: 23  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Kathryn A. Seese, Immunex Corporation  
: STREET: 51 University Street  
: CITY: Seattle  
: STATE: Washington  
: COUNTRY: USA  
: ZIP: 98101  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: Apple Macintosh  
: OPERATING SYSTEM: Apple 7.1  
: SOFTWARE: Microsoft Word, Version 5.1a  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/225,989  
: FILING DATE: 12 APRIL 1994  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/966,775  
: FILING DATE: 27-OCT-1992  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 907,224  
: FILING DATE: 01-JUL-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 899,660  
: FILING DATE: 15-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 892,459  
: FILING DATE: 02-JUN-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 889,717  
: FILING DATE: 26-MAY-1992  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Seese, Kathryn A.  
: REGISTRATION NUMBER: 32,172  
: REFERENCE/DOCKET NUMBER: 2804-E  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (206)587-0430  
: TELEFAX: (206)233-0644

TELEX: 756822  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
IMMEDIATE SOURCE:  
CLONE: 5' PCR Primer  
US-08-225-989-9

Query Match: 76.0%; Score 15.2; DB 1; Length 39;  
Best Local Similarity 85.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTTG 20  
||| ||||| ||||| |||  
Db 31 CGAGGAGGAGCGCATGTTG 12

RESULT 2  
US-08-570-923-9/c  
Sequence 9, Application US/08570923  
Patent No. 5677430  
GENERAL INFORMATION:  
APPLICANT: Goodwin, Raymond G.  
APPLICANT: Smith, Craig A.  
APPLICANT: Armitage, Richard J.  
APPLICANT: Gruss, Hans-Jurgen  
TITLE OF INVENTION: No. 5677430a1 Cytokine That Binds CD30  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kathryn A. Seese, Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/570/923  
FILING DATE: 12-DEC-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,989  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: US 07/966,775  
FILING DATE: 27-OCT-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 907,224  
FILING DATE: 01-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,660  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 892,459  
FILING DATE: 02-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 889,717  
FILING DATE: 26-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2804-E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)233-0644

TELEX: 756822  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
IMMEDIATE SOURCE:  
CLONE: 5' PCR Primer  
US-08-570-923-9

Query Match: 76.0%; Score 15.2; DB 1; Length 39;  
Best Local Similarity 85.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTTG 20  
||| ||||| ||||| |||  
Db 31 CGAGGAGGAGCGCATGTTG 12

RESULT 3  
US-08-580-014-9/c  
Sequence 9, Application US/08580014  
Patent No. 5753203  
GENERAL INFORMATION:  
APPLICANT: Goodwin, Raymond G.  
APPLICANT: Smith, Craig A.  
APPLICANT: Armitage, Richard J.  
APPLICANT: Gruss, Hans-Jurgen  
TITLE OF INVENTION: No. 5753203e1 Cytokine That Binds CD30  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kathryn A. Seese, Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/580,014  
FILING DATE: 20-DEC-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,989  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: US 07/966,775  
FILING DATE: 27-OCT-1992  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 907,224  
FILING DATE: 01-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,660  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 892,459  
FILING DATE: 02-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 889,717  
FILING DATE: 26-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2804-E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)233-0644

TELEX: 756822  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
IMMEDIATE SOURCE:  
CLONE: 5' PCR Primer  
US-08-580-014-9

Query Match 76.0%; Score 15.2; DB 1; Length 39;  
Best Local Similarity 85.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||  
Db 31 CGAGGAGGACGCGCATGTG 12

RESULT 4:  
US-09-079-785-9/c  
Sequence 9, Application US/09079785  
Patent No. 6143869  
GENERAL INFORMATION:  
APPLICANT: Goodwin, Raymond G.  
APPLICANT: Smith, Craig A.  
APPLICANT: Armitage, Richard J.  
TITLE OF INVENTION: Gruss, Hans-Jurgen  
FILE REFERENCE: No. 6143869el Cytokine That Binds CD30  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kathryn A. Seese, Immunex Corporation  
STREET: 51 University Street  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Apple 7.1  
SOFTWARE: Microsoft Word, Version 5.1a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/079,785  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/225,989  
FILING DATE: 12 APRIL 1994  
APPLICATION NUMBER: US 07/966,775  
FILING DATE: 27-OCT-1992  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 907,224  
FILING DATE: 01-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 899,660  
FILING DATE: 15-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 892,459  
FILING DATE: 02-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 889,717  
FILING DATE: 26-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Seese, Kathryn A.  
REGISTRATION NUMBER: 32,172  
REFERENCE/DOCKET NUMBER: 2804-E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206)587-0430  
TELEFAX: (206)233-0644

TELEX: 756822  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
IMMEDIATE SOURCE:  
CLONE: 5' PCR Primer  
US-09-079-785-9

Query Match 76.0%; Score 15.2; DB 3; Length 39;  
Best Local Similarity 85.0%; Pred. No. 63;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||  
Db 31 CGAGGAGGACGCGCATGTG 12

RESULT 5  
US-09-587-700-1  
Sequence 1, Application US/09587700  
Patent No. 6291666  
GENERAL INFORMATION:  
APPLICANT: Puthigae, Sathish  
APPLICANT: Skadsen, Ronald W.  
TITLE OF INVENTION: Spike Tissue Specific Promoter  
FILE REFERENCE: 960296.97231  
CURRENT APPLICATION NUMBER: US/09/587,700  
CURRENT FILING DATE: 2000-06-05  
PRIOR APPLICATION NUMBER: 60\204,040  
PRIOR FILING DATE: 2000-05-12  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 1  
LENGTH: 1434  
TYPE: DNA  
ORGANISM: Hordeum vulgare  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (411)  
NAME/KEY: promoter  
LOCATION: (1187)..(1237)  
NAME/KEY: misc\_feature  
LOCATION: (939)..(1014)  
OTHER INFORMATION: Shares partial identity with the promoter region  
of the rice anther-specific gene (RTS2)  
US-09-587-700-1

Query Match 76.0%; Score 15.2; DB 4; Length 1434;  
Best Local Similarity 85.0%; Pred. No. 82;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||  
Db 253 CGCTGACGAGCGCTGTG 272

RESULT 6  
US-09-221-017B-545/c  
Sequence 545, Application US/09221017B  
Patent No. 6444799  
GENERAL INFORMATION:  
APPLICANT: Ross, Bruce C.  
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
NUMBER OF SEQUENCES: 1120  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA

COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P1182  
FILING DATE: 31-DEC-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P1546  
FILING DATE: 30-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P2911  
FILING DATE: 09-APR-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU98/01023  
FILING DATE: 10-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Monroy, Gladys H  
REGISTRATION NUMBER: 32,430  
REFERENCE/DOCKET NUMBER: 27340-20021.00  
TELEPHONE: 650-813-5600  
TELEFAX: 650-494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 545:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5633 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: UNKNOWN  
ORIGINAL SOURCE:  
ORGANISM: PORYPHYROMONAS GINGIVALIS  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1...5633  
US-09-221-017B-545

Query Match 76.0%; Score 15.2; DB 4; Length 5633;  
Best Local Similarity 85.0%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTG 20  
||||| ||||| ||||| ||||| |||||  
Db 3685 CGATACGGTGCCTGTG 3666

RESULT 7  
US-09-215-694-19  
Sequence 19, Application US/09215694B  
Patent No. 6391583  
GENERAL INFORMATION:  
APPLICANT: Wisconsin Alumni Research Foundation  
APPLICANT: Hutchinson, Charles R.  
APPLICANT: Kennedy, Jonathan n.m.i  
APPLICANT: Park, Cheonseok n.m.i  
TITLE OF INVENTION: METHOD OF PRODUCING ANTIHYPERCHOLESTEROLEMIC AGENTS  
FILE REFERENCE: 960296.95718  
CURRENT APPLICATION NUMBER: US/09/215,694B  
CURRENT FILING DATE: 1999-12-18  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 19  
LENGTH: 31328

TYPE: DNA  
ORGANISM: Aspergillus terreus  
US-09-215-694-19

Query Match 76.0%; Score 15.2; DB 4; Length 31328;  
Best Local Similarity 85.0%; Pred. No. 1e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTG 20  
||||| ||||| ||||| ||||| |||||  
Db 28294 CGAAGATGACGCGCTATTG 28313

RESULT 8  
US-09-103-840A-2  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 76.0%; Score 15.2; DB 4; Length 4403765;  
Best Local Similarity 85.0%; Pred. No. 61;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTG 20  
||||| ||||| ||||| ||||| |||||  
Db 2233178 CGATGACGCGCGCGTGTG 2233197

RESULT 9  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
TITLE OF INVENTION: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37RV  
US-09-103-840A-1

Query Match 76.0%; Score 15.2; DB 4; Length 4411529;  
Best Local Similarity 85.0%; Pred. No. 61;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
|||||...|||...|||  
Db 2235879 CGATGAGGACGCGCGTGTG 2235898

## RESULT 10

US-08-635-121-1  
; Sequence 1, Application US/08635121  
; Patent No. 5910442

## GENERAL INFORMATION:

APPLICANT: Geiman, Irwin H.  
TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10112-0228

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSEQ Version 1.5

CURRENT APPLICATION DATA: US/08/635,121

FILING DATE: 19-APRIL-1996

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Richard S

REGISTRATION NUMBER: 26,154

REFERENCE/DOCKET NUMBER: A30558 - 165/33603

TELEPHONE: 212-408-2558

TELEFAX: 212-765-2519

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5134 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE:

ORIGINAL SOURCE:

US-08-635-121-1

Query Match 74.0%; Score 14.8; DB 2; Length 5134;

Best Local Similarity 88.9%; Pred. No. 1.4e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATGAGGACGCGCTGTG 20

|||||...|||...|||

Db 1476 ATGAGGACGCGCTGTG 1493

## RESULT 11

US-08-125-468-1/c

; Sequence 1, Application US/08125468

; Patent No. 5589385

GENERAL INFORMATION:

APPLICANT: Ryan, Michael J.

APPLICANT: Lotvin, Jason A.

APPLICANT: Strathy, Nancy

APPLICANT: Fantini, Susan E.

TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosms  
TITLE OF INVENTION: useful therein  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: American Cyanamid Company  
STREET: One Cyanamid Plaza  
CITY: Wayne  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/125,468

FILING DATE: 22-SEP-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Tsevdos, Estelle J

REGISTRATION NUMBER: 31,145

REFERENCE/DOCKET NUMBER: 31,255-02

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201)831-3241

TELEFAX: (201)831-3305

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 30001 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-125-468-1

Query Match 74.0%; Score 14.8; DB 1; Length 30001;

Best Local Similarity 88.9%; Pred. No. 1.6e+02;

Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGT 18

|||||...|||...|||

Db 26264 CGATGAGGACGCGTCCGT 26247

## RESULT 12

US-08-474-933-1/c

; Sequence 1, Application US/08474933

; Patent No. 5866410

GENERAL INFORMATION:

APPLICANT: Ryan, Michael J.

APPLICANT: Lotvin, Jason A.

APPLICANT: Strathy, Nancy

APPLICANT: Fantini, Susan E.

TITLE OF INVENTION: Cloning of the biosynthetic pathway for

TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosms

TITLE OF INVENTION: useful therein

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: USA

ZIP: 07470

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,933

FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/125,468  
FILING DATE: 22-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsevdos, Estelle J  
REGISTRATION NUMBER: 31,145  
REFERENCE/DOCKET NUMBER: 31,255-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)831-3241  
TELEFAX: (201)831-3305  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-474-933-1

Query Match 74.0%; Score 14.8; DB 2; Length 30001;  
Best Local Similarity 88.9%; Pred. No. 1.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGATGAGGACGGCGCTGT 18  
|||||  
Db 26264 CGATGAGGACGGCGT 26247

RESULT 13  
US-08-529-600D-1  
Sequence 1, Application US/08529600D  
Patent No. 5861285  
GENERAL INFORMATION:  
APPLICANT: Tadashi MATSUNAGA  
TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,600D  
FILING DATE: 18-SEP-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-248700  
FILING DATE: 16-SEP-1994  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1302 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1302  
US-08-529-600D-1

Query Match 72.0%; Score 14.4; DB 2; Length 1302;  
Best Local Similarity 93.8%; Pred. No. 2e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GATGAGGACGGCGCTG 17  
|||||  
Db 69 GATGATGACGGCGCTG 84

RESULT 14  
US-08-973-275-5  
Sequence 5, Application US/08973275B  
Patent No. 5958706  
GENERAL INFORMATION:  
APPLICANT: MATSUNAGA, Tadashi  
APPLICANT: KAMIYA, Shinji  
APPLICANT: NAMBA, Kenryo  
TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS  
TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: MATSUNAGA  
CURRENT APPLICATION NUMBER: US/08/973,275B  
CURRENT FILING DATE: 1998-02-09  
EARLIER APPLICATION NUMBER: PCT/JP97/01043  
EARLIER FILING DATE: 1997-03-27  
EARLIER APPLICATION NUMBER: JP 8-97536  
EARLIER FILING DATE: 1996-03-28  
EARLIER APPLICATION NUMBER: JP 8-146833  
EARLIER FILING DATE: 1996-05-16  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 5  
LENGTH: 1302  
TYPE: DNA  
ORGANISM: Magnetospirillum AMB-1  
FEATURE:  
NAME/KEY: gene  
LOCATION: (1)..(1302)  
OTHER INFORMATION: Genomic DNA sequence encoding mag A gene.  
US-08-973-275-5

Query Match 72.0%; Score 14.4; DB 2; Length 1302;  
Best Local Similarity 93.8%; Pred. No. 2e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GATGAGGACGGCGCTG 17  
|||||  
Db 69 GATGATGACGGCGCTG 84

RESULT 15  
US-09-122-632-1  
Sequence 1, Application US/09122632  
Patent No. 6033878  
GENERAL INFORMATION:  
APPLICANT: Tadashi MATSUNAGA  
TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND  
TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.  
STREET: 1100 NEW YORK AVENUE, N.W.  
CITY: WASHINGTON  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3918  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/122,632  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,600

;; FILING DATE: 18-SEP-1995  
;; APPLICATION NUMBER: JP 6-248700  
;; FILING DATE: 16-SEP-1994  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1302 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: DNA (genomic)  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: 1..1302  
US-09-122-632-1

Query Match 72.0%; Score 14.4; DB 3; Length 1302;  
Best Local Similarity 93.8%; Pred. No. 2e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GATGAGGACGCGCCTG 17  
|||||  
Db 69 GATGATGACGCGCCTG 84

Search completed: July 8, 2003, 09:31:59  
Job time : 39.6598 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 115.512 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-13

Perfect score: 20

Sequence: 1 cgatgagcgcgcctgttg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*\*
- 2: /cgn2\_6/ptodata/2/pubpna/PTC\_NEW\_PUB.seq:\*\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*\*
- 5: /cgn2\_6/ptodata/2/pubpna/PTC\_US\_PUBCOMB.seq:\*\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	20	100.0	2893	9 US-10-017-273A-4	Sequence 4, Appli
C 2	20	100.0	2893	10 US-09-905-846-1	Sequence 1, Appli
C 3	20	100.0	2975	9 US-10-017-273A-5	Sequence 5, Appli
C 4	20	100.0	2975	10 US-09-905-846-5	Sequence 21, Appli
5	18.4	92.0	300	9 US-09-990-099-21	Sequence 21, Appli
6	15.8	79.0	351	10 US-09-960-352-14675	Sequence 14675, A
7	15.8	79.0	1860	10 US-09-974-300-2780	Sequence 2780, Ap
8	15.8	79.0	1908	10 US-09-974-300-2714	Sequence 2714, Ap
9	15.8	79.0	10144	10 US-09-880-107-2168	Sequence 2168, Ap
10	15.4	77.0	447	10 US-09-983-965-3946	Sequence 3946, Ap
11	15.2	76.0	207	10 US-09-864-761-20314	Sequence 20314, A
12	15.2	76.0	437	10 US-09-864-761-3544	Sequence 3544, Ap
13	15.2	76.0	639	9 US-09-938-842A-126	Sequence 126, App
14	15.2	76.0	639	9 US-09-938-842A-481	Sequence 481, App
15	15.2	76.0	814	9 US-09-989-643-127	Sequence 127, App
16	15.2	76.0	846	9 US-10-156-761-2628	Sequence 2628, Ap
17	15.2	76.0	891	9 US-09-989-643-160	Sequence 160, App
18	15.2	76.0	954	9 US-10-156-761-2747	Sequence 2747, Ap
19	15.2	76.0	968	10 US-09-770-445-293	Sequence 293, App

C 20	15.2	76.0	996	10	US-09-815-242-7811	Sequence 7811, Ap
C 21	15.2	76.0	1023	9	US-10-156-761-7102	Sequence 7102, Ap
C 22	15.2	76.0	1029	9	US-10-103-313-216	Sequence 216, App
C 23	15.2	76.0	1397	9	US-10-103-313-69	Sequence 69, Appli
C 24	15.2	76.0	1479	9	US-09-738-626-1393	Sequence 1393, Ap
C 25	15.2	76.0	1625	10	US-09-796-766-17	Sequence 17, Appli
C 26	15.2	76.0	4125	9	US-10-121-988-76	Sequence 76, Appli
C 27	15.2	76.0	4125	9	US-10-121-988-149	Sequence 149, App
C 28	15.2	76.0	4149	9	US-10-121-988-84	Sequence 84, Appli
C 29	15.2	76.0	4509	9	US-10-124-800-5	Sequence 5, Appli
C 30	15.2	76.0	4512	9	US-10-124-800-27	Sequence 27, Appli
C 31	15.2	76.0	4767	9	US-10-331-061-76	Sequence 76, Appli
C 32	15.2	76.0	7569	9	US-10-072-349-333	Sequence 333, App
C 33	15.2	76.0	7569	9	US-10-072-349-334	Sequence 334, App
C 34	15.2	76.0	7569	10	US-09-764-855-333	Sequence 333, App
C 35	15.2	76.0	7569	10	US-09-764-855-334	Sequence 334, App
C 36	15.2	76.0	3309400	9	US-09-738-626-1	Sequence 1, Appli
C 37	15.2	76.0	9025608	9	US-10-156-761-1	Sequence 1, Appli
C 38	15	75.0	951	9	US-09-738-626-1362	Sequence 1362, Ap
C 39	15	75.0	1215	9	US-09-938-842A-58	Sequence 58, Appli
C 40	14.8	74.0	3585	9	US-10-287-290-2	Sequence 2, Appli
C 41	14.8	74.0	3617	9	US-10-287-290-8	Sequence 8, Appli
C 42	14.8	74.0	5134	9	US-09-902-432-1	Sequence 1, Appli
C 43	14.8	74.0	1830121	9	US-10-329-960-1	Sequence 1, Appli
C 44	14.8	74.0	9025608	9	US-10-156-761-1	Sequence 1, Appli
C 45	14.4	72.0	25	9	US-10-215-112-4023	Sequence 4023, Ap

## ALIGNMENTS

RESULT 1  
US-10-017-273A-4/c  
; Sequence 4, Application US/10017273A  
; Publication No. US20030119714A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Inc.  
; APPLICANT: Naylor, Alasdair M.  
; APPLICANT: Van Der Graaf, Pieter H  
; APPLICANT: Wayman, Christopher P.  
; TITLE OF INVENTION: Treatment of Male Sexual Dysfunction  
; FILE REFERENCE: PC22013  
; CURRENT APPLICATION NUMBER: US/10/017,273A  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/265,358  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: GB 0030647.2  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: GB 0108730.3  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: GB 0120679.6  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 09/905,846  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/291,722  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: US 09/895,367  
; PRIOR FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 2893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-017-273A-4

Query Match 100.0%; Score 20; DB 9; Length 2893;  
Best Local Similarity 100.0%; Pred No. 0.74;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGCAGCGCCTGTG 20  
|||||

DB 884 CGATGAGCAGCGCCTGTG 865

RESULT 2  
US-09-905-846-1/c  
; Sequence 1, Application US/09905846  
; Patent No. US20020102707A1  
; GENERAL INFORMATION:  
; APPLICANT: Ian Dennis Harrow  
; APPLICANT: Peter Stacey  
; APPLICANT: Rogerick Thomas Walsh  
; TITLE OF INVENTION: Compounds for the treatment of sexual dysfunction  
; FILE REFERENCE: PCS10926APME  
; CURRENT APPLICATION NUMBER: US/09/905,846  
; CURRENT FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: 0017387.2  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/220,908  
; PRIOR FILING DATE: 2000-07-26  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-905-846-1

Query Match 100.0%; Score 20; DB 10; Length 2893;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
Db 884 CGATGAGGACGCGCTGTG 865

RESULT 3  
US-10-017-273A-5/c  
; Sequence 5, Application US/10017273A  
; Publication No. US20030119714A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Inc.  
; APPLICANT: Naylor, Alasdair M.  
; APPLICANT: Van Der Graaf, Pieter H  
; APPLICANT: Wayman, Christopher P.  
; TITLE OF INVENTION: Treatment of Male Sexual Dysfunction  
; FILE REFERENCE: PC22013  
; CURRENT APPLICATION NUMBER: US/10/017,273A  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/265,358  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: GB 0030647.2  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: GB 0108730.3  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: GB 0120679.6  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 09/905,846  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/291,722  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: US 09/895,367  
; PRIOR FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 2975  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-017-273A-5

Query Match 100.0%; Score 20; DB 9; Length 2975;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
Db 949 CGATGAGGACGCGCTGTG 930

RESULT 4  
US-09-905-846-5/c  
; Sequence 5, Application US/09905846  
; Patent No. US20020102707A1  
; GENERAL INFORMATION:  
; APPLICANT: Ian Dennis Harrow  
; APPLICANT: Peter Stacey  
; APPLICANT: Rogerick Thomas Walsh  
; TITLE OF INVENTION: Compounds for the treatment of sexual dysfunction  
; FILE REFERENCE: PCS10926APME  
; CURRENT APPLICATION NUMBER: US/09/905,846  
; CURRENT FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: 0017387.2  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/220,908  
; PRIOR FILING DATE: 2000-07-26  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 2975  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-905-846-5

Query Match 100.0%; Score 20; DB 10; Length 2975;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
Db 949 CGATGAGGACGCGCTGTG 930

RESULT 5  
US-09-990-099-21  
; Sequence 21, Application US/09990099  
; Publication No. US20030119094A1  
; GENERAL INFORMATION:  
; APPLICANT: Lesley, Scott A.  
; APPLICANT: Knuth, Mark  
; TITLE OF INVENTION: Solubility Reporter Gene Constructs  
; FILE REFERENCE: P0012US20  
; CURRENT APPLICATION NUMBER: US/09/990,099  
; CURRENT FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US 09/721,340  
; PRIOR FILING DATE: 2000-11-21  
; PRIOR APPLICATION NUMBER: US 60/324,833  
; PRIOR FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 21  
; LENGTH: 300  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
US-09-990-099-21

Query Match 92.0%; Score 18.4; DB 9; Length 300;  
Best Local Similarity 95.0%; Pred. No. 5.1;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
Db 210 CGATGAGGACGCGCTGTG 229

RESULT 6  
US-09-960-352-14675

```
; Sequence 14675, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14675
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 63-LIB188-003-Q1-E1-H4
US-09-960-352-14675

Query Match          79.0%; Score 15.8; DB 10; Length 351;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GATGAGGAGCGCCCTGTTG 20
Db      85 GATGAGGAGCGCCCTGTAG 103

RESULT 7
US-09-974-300-2780
; Sequence 2780, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-05
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2780
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2780

Query Match          79.0%; Score 15.8; DB 10; Length 1860;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGATGAGGAGCGCGCTGTT 19
Db      1158 CCATGAGGAGCGCTGTT 1176

RESULT 8
US-09-974-300-2714
; Sequence 2714, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
```

```
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2714
; LENGTH: 1908
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2714

Query Match          79.0%; Score 15.8; DB 10; Length 1908;
Best Local Similarity 89.5%; Pred. No. 93;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGATGAGGAGCGCGCTGTT 19
Db      1223 CCATGAGGAGCGCTGTT 1241

RESULT 9
US-09-880-107-2168/c
; Sequence 2168, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2168
; LENGTH: 10144
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 L17131
; NAME/KEY: unsure
; LOCATION: (1)..(10144)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-2168

Query Match          79.0%; Score 15.8; DB 10; Length 10144;
Best Local Similarity 89.5%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CGATGAGGAGCGCGCTGTT 19
Db      1013 CGCTGAGGAGCGCGCTGCT 995

RESULT 10
US-09-983-965-3946
; Sequence 3946, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
; FILE REFERENCE: 37-21(10297)C
```

;; CURRENT APPLICATION NUMBER: US/09/983,965  
;; CURRENT FILING DATE: 2001-10-26  
;; PRIOR APPLICATION NUMBER: US 09/465,231  
;; PRIOR FILING DATE: 1999-12-15  
;; PRIOR APPLICATION NUMBER: US 60/113,678  
;; PRIOR FILING DATE: 1998-12-17  
;; NUMBER OF SEQ ID NOS: 5912  
;; SEQ ID NO 3946  
;; LENGTH: 447  
;; TYPE: DNA  
;; ORGANISM: Bos taurus  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: 52-LIB3058-005-Q1-K1-E8  
US-09-983-965-3946

Query Match 77.0%; Score 15.4; DB 10; Length 447;  
Best Local Similarity 94.1%; Pred. No. 1.6e-02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGAGGAGCGCGCTGTG 20  
||||| |||||||  
DB 167 TGAGGAGTCGCGCTGTG 183

## RESULT 11

US-09-864-761-20314  
;; Sequence 20314, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng

;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761

;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203

;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 20314  
;; LENGTH: 207  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL021154.1  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.6  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.4  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.2  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.8  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.9  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.8  
;; OTHER INFORMATION: SWISSPROT HIT: P75334, EVALUATE 8.30e-02  
;; OTHER INFORMATION: NT HIT: M26533.1, EVALUATE 1.90e+00  
US-09-864-761-20314

Query Match 76.0%; Score 15.2; DB 10; Length 207;  
Best Local Similarity 85.0%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGAGCGCGCTGTG 20  
|| ||||| |||||  
DB 103 CGGTGAGGAGCGCAGCTGTG 122

## RESULT 12

US-09-864-761-3544  
;; Sequence 3544, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.

;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL  
;; FILE REFERENCE: Aecomica-X-1  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 3544  
LENGTH: 437  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL021154.1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.6  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 6.4  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.2  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.8  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.8  
US-09-864-761-3544

Query Match 76.0%; Score 15.2; DB 10; Length 437;  
Best Local Similarity 85.0%; Pred. No. 2e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||||  
Db 315 CGGTGAGGACGCGCTGTG 334

RESULT 13  
US-09-938-842A-126/C  
Sequence 126, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRI1300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 126  
LENGTH: 639  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-126

Query Match 76.0%; Score 15.2; DB 9; Length 639;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||||  
Db 61 CGATGAGGACTCTCTGTG 42

## RESULT 14

US-09-938-842A-481/C  
Sequence 481, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Kreps, Joel  
APPLICANT: Wang, Xun  
APPLICANT: Zhu, Tong  
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
FILE REFERENCE: SCRI1300-3  
CURRENT APPLICATION NUMBER: US/09/938,842A  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: US 60/227,866  
PRIOR FILING DATE: 2000-08-24  
PRIOR APPLICATION NUMBER: US 60/264,647  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/300,111  
PRIOR FILING DATE: 2001-06-22  
NUMBER OF SEQ ID NOS: 5379  
SEQ ID NO 481  
LENGTH: 639  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-938-842A-481

Query Match 76.0%; Score 15.2; DB 9; Length 639;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||||  
Db 61 CGATGAGGACTCTCTGTG 42

## RESULT 15

US-09-989-643-127  
Sequence 127, Application US/09989643  
Publication No. US20030049636A1  
GENERAL INFORMATION:  
APPLICANT: Bergeron, Michel G.  
APPLICANT: Picard, Francois J.  
APPLICANT: Ouellette, Marc  
APPLICANT: Roy, Paul H.  
TITLE OF INVENTION: Species-Specific, Genus-Specific and Universal DNA  
TITLE OF INVENTION: Probes and Amplification Primers to Rapidly Detect and  
TITLE OF INVENTION: Identify Common Bacterial and Fungal Pathogens and  
TITLE OF INVENTION: Associated Antibiotic Resistance Genes from  
FILE REFERENCE: 12287.29  
CURRENT APPLICATION NUMBER: US/09/989,643  
CURRENT FILING DATE: 2001-11-20  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/297,539  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-05-09  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/743,637  
PRIOR FILING DATE: EARLIER FILING DATE: 1996-11-04  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 127  
LENGTH: 814  
TYPE: DNA  
ORGANISM: Corynebacterium genitalium  
US-09-989-643-127

Query Match 76.0%; Score 15.2; DB 9; Length 814;  
Best Local Similarity 85.0%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20  
||| ||||| ||||| |||||  
Db 190 CGACGAGGAGCAGCCTGTG 209

Search completed: July 9, 2003, 02:22:19  
Job time : 126.562 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1013.95 seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-13  
Perfect score: 20  
Sequence: 1 cgatgagacgcctgttg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpi:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_estfun:\*  
15: em\_estom:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	18.4	92.0	141	14	BQ668192
C 2	18.4	92.0	260	10	AW261276
C 3	18.4	92.0	305	9	AI691198
C 4	18.4	92.0	326	10	AW600514
C 5	18.4	92.0	349	10	AW562792
C 6	18.4	92.0	364	10	AW562970

C 7	18.4	92.0	422	14	BQ703364
C 8	18.4	92.0	436	12	BG837021
C 9	18.4	92.0	444	10	AW202460
C 10	18.4	92.0	464	14	BQ635402
C 11	18.4	92.0	466	14	BQ037446
C 12	18.4	92.0	481	13	B1431034
C 13	18.4	92.0	489	10	AW313130
C 14	18.4	92.0	494	14	BQ744736
C 15	18.4	92.0	498	14	BQ294319
C 16	18.4	92.0	506	9	AI668369
C 17	18.4	92.0	528	9	AI834379
C 18	18.4	92.0	537	9	AI665885
C 19	18.4	92.0	542	9	AI737277
C 20	18.4	92.0	547	10	AW562522
C 21	18.4	92.0	549	10	AW067236
C 22	18.4	92.0	550	10	AW181220
C 23	18.4	92.0	552	9	AI691236
C 24	18.4	92.0	555	10	AW352676
C 25	18.4	92.0	556	12	BG833132
C 26	18.4	92.0	563	12	BG833131
C 27	18.4	92.0	565	10	AW062014
C 28	18.4	92.0	572	9	AI692074
C 29	18.4	92.0	583	10	AW191088
C 30	18.4	92.0	585	14	BQ668198
C 31	18.4	92.0	587	9	AI691181
C 32	18.4	92.0	588	10	AW256209
C 33	18.4	92.0	600	13	B1989738
C 34	18.4	92.0	608	9	AI691299
C 35	18.4	92.0	621	9	AI691802
C 36	18.4	92.0	642	10	BE553003
C 37	18.4	92.0	675	14	BQ529205
C 38	18.4	92.0	821	12	BG837038
C 39	18.4	92.0	2534	11	AY103538
C 40	17.4	87.0	341	10	BB701995
C 41	17.4	87.0	430	10	BB780245
C 42	17.4	87.0	464	17	AZ180792
C 43	16.8	84.0	236	12	BF991788
C 44	16.8	84.0	251	13	BM645820
C 45	16.8	84.0	285	10	AW061907

#### ALIGNMENTS

RESULT 1:  
LOCUS BQ668192/c 141 bp mRNA linear EST 15-JUL-2002  
DEFINITION 946103H11.x1 946 --tassel primordium prepared by Schmidt lab Zea mays CDNA, mRNA sequence.  
ACCESSION BQ668192  
VERSION BQ668192.1 GI:21809874  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 141)  
AUTHORS Walbot V.  
TITLE Zea mays from various cDNA libraries sequenced at Stanford University  
JOURNAL Unpublished (1999)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 946103 row: H column: 11.  
FEATURES  
Location/Qualifiers  
1..141  
/organism="Zea mays"

BQ703364 946108G07  
BG837021 2m08\_1041  
AW202460 687030G10  
BQ635402 109107080  
BU037446 946139A11  
B1431034 949064D12  
AW313130 660021C10  
BQ744736 946108G07  
BQ294319 109102880  
AI668369 605031D07  
AI834379 606067C11  
AI665885 606003B12  
AI737277 606038C12  
AW562522 660065H10  
AW067236 683019E03  
AW181220 687015C05  
AI691236 606025H12  
AW352676 660034B08  
BG833132 951001C03  
BG833131 951001C03  
AW062014 687010E01  
AI692074 606011E10  
AW191088 687017D02  
BQ668198 946104A03  
AI691181 606025B10  
AW256209 687075E03  
B1989738 4044-75 M  
AI691299 606028H11  
AI691802 606009C11  
BE553003 946088D01  
BQ329205 1091056A0  
BG837038 2m08\_10H0  
AY103538 Zea mays  
BB701995 BB701995  
BB780245 BB780245  
AZ180792 SP\_0181\_A  
BF991788 MRO-GN013  
BM645820 170006873  
AW061907 687008H11

```

/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassels primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/notes="Organ: tassels; Vector: HybridAP; Site_1: EcoRI;
Site_2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
BASE COUNT      31 a      36 c      47 g      27 t
ORIGIN

Query Match.      92.0%; Score 18.4; DB 14; Length 141;
Best Local Similarity 95.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGATGAGGACGCGCTGTG 20
|||||  |||||  |||||  |||||
Db 21 CGATGAGGACGCGCTGTG 2

RESULT 2
AW261276/c
LOCUS      687076D03.y1 687 - Early embryo from Delaware Zea mays cDNA, mRNA
sequence.
ACCESSION  AW261276.1 GI:6637004
VERSION     AW261276
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 260)
AUTHORS     Walbot,V.
TITLES      Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 687076 row: D column: 03.
FEATURES
source
Location/Qualifiers
1..260
/organism="Zea mays"
/cultivar="Illinois High Oil"
/db_xref="taxon:4577"
/clone_lib="687 - Early embryo from Delaware"
/tissue_type="embryo"
/dev_stage="14, 21, 28, and 35 days after pollination"
/lab_host="E. coli SOLR"
/notes="Organ: embryo; Vector: pBluescript SK; Site_1: XhoI
; Site_2: EcoRI; Library was prepared by Statagene using
the Uni-ZAP XR system (Stratagene BN937328-12). Clones
were picked by a Q-bot after blue/white selection
(ampicillin resistance - use 100 micrograms/microliter).
Developed from a pool of equal amounts of RNA from
developing embryos sampled at 14, 21, 28 and 35 days after
pollination of the Illinois High Oil Zea Strain Cycle
90. This closed strain has been selected for high oil
concentration for 90 generations and originates from the
1890s era open pollinated variety Burr's White"
BASE COUNT      82 a      68 c      50 g      60 t
ORIGIN

```

```

Query Match      92.0%; Score 18.4; DB 10; Length 260;
Best Local Similarity 95.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGATGAGGACGCGCTGTG 20
|||||  |||||  |||||  |||||
Db 247 CGATGAGGACGCGCTGTG 228

RESULT 3
AI691198/c
LOCUS      606025D10.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION  Zea mays cDNA, mRNA sequence.
ACCESSION  AI691198.1 GI:4938793
VERSION     AI691198.1
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 305)
AUTHORS     Walbot,V.
TITLES      Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606025 row: D column: 10.
FEATURES
source
Location/Qualifiers
1..305
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt
lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/notes="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcoRI
; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"
BASE COUNT      94 a      75 c      54 g      82 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 9; Length 305;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CGATGAGGACGCGCTGTG 20
|||||  |||||  |||||  |||||
Db 301 CGATGAGGACGCGCTGTG 282

RESULT 4
AW600514/c
LOCUS      660070D05.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION  mRNA sequence.
ACCESSION  AW600514.1 GI:7304575
VERSION     AW600514
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

```

```

clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 326)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
COMMENT University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660070 row: D column: 05.
FEATURES
SOURCE 1
Location/Qualifiers
1..326
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XL0LR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 84 a 90 c 86 g 66 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 326;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
|||||
Db 206 CGATGAGGACGCGCTGTTG 187

RESULT 5
AW562792 349 bp mRNA linear EST 10-MAR-2000
LOCUS 660065H10.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AW562792
VERSION AW562792.1 GI:7216670
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 349)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
COMMENT University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660065 row: H column: 10.
FEATURES
SOURCE 1
Location/Qualifiers
1..349
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"

clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 349)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
COMMENT University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660070 row: D column: 05.
FEATURES
SOURCE 1
Location/Qualifiers
1..326
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XL0LR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 84 a 90 c 86 g 66 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 326;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
|||||
Db 206 CGATGAGGACGCGCTGTTG 187

RESULT 5
AW562792 349 bp mRNA linear EST 10-MAR-2000
LOCUS 660065H10.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AW562792
VERSION AW562792.1 GI:7216670
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 349)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
COMMENT University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660070 row: D column: 05.
FEATURES
SOURCE 1
Location/Qualifiers
1..326
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XL0LR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 84 a 90 c 86 g 66 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 349;
Best Local Similarity 95.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
|||||
Db 232 CGATGAGGACGCGCTGTTG 251

RESULT 6
AW562970 364 bp mRNA linear EST 10-MAR-2000
LOCUS 660070D05.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AW562970
VERSION AW562970.1 GI:7216848
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
AUTHORS 1 (bases 1 to 364)
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
COMMENT University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660070 row: D column: 05.
FEATURES
SOURCE 1
Location/Qualifiers
1..364
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XL0LR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."
BASE COUNT 88 a 87 c 89 g 100 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 10; Length 364;
Best Local Similarity 95.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTTG 20
|||||
Db 105 CGATGAGGACGCGCTGTTG 124

RESULT 7
BQ703364/c 422 bp mRNA linear EST 15-JUL-2002
LOCUS 946108G07.xl 946 - tassell primordium prepared by schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.

```

```

ACCESSION BQ703364
VERSION BQ703364.1 GI:21828680
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 422)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford.
JOURNAL
COMMENT Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946108 row: G column: 07.
Location/Qualifiers
1. 422
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946" tassel primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/Note="Organ: tassels; Vector: HybridZAP; Site_1: EcoRI;
Site_2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 kb with a 1 kb average."
BASE COUNT 117 a 111 c 105 g 89 t
ORIGIN
Query Match 92.0%; Score 18.4; DB 14; Length 422;
Best Local Similarity 95.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20
|||||
Db 249 CGATGAGGACGCGCTGTG 230

RESULT 8
BG837021
LOCUS Zm08_10d11.A 436 bp mRNA linear EST 25-MAY-2001
DEFINITION Zm08_AAFECORC_Fusarium_graminearum_inoculated_corn_ear Zea mays
CDNA clone Zm08_10d11, mRNA sequence.
ACCESSION BG837021
VERSION BG837021.1 GI:14203344
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 436)
AUTHORS Harris,L.J., Balcerzak,M., Allard,S., Saparno,A., Couroux,P., De
Moors,A., Hattori,J.I., Ouellet,T., Robert,L.S., Singh,J.A, Sprott
D. and Tinker,N.A.
TITLE Expressed Sequence Tags from Developing Maize Kernels Six Days
after Silk Channel Inoculation with Fusarium graminearum
JOURNAL Unpublished (2001)
COMMENT Contact: Harris, Linda J.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada

Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,
CANADA
Tel: (613) 759-1314
Fax: (613) 759-6566
Email: harris@em.agr.ca.
Location/Qualifiers
1. 436
/organism="Zea mays"
/cultivar="CO430"
/db_xref="taxon:4577"
/clone_lib="Zm08_10d11"
/clone_lib="Zm08_AAFECORC_Fusarium_graminearum_inoculate
d_corn_ear"
/tissue_type="developing kernels (sibcrossed)"
/dev_stage="10-11 days post-silk emergence"
/Note="Vector: Bluescript SK+/XhoI-EcoRI; Site_1: EcoRI;
Site_2: XhoI; Field-grown maize ears were silk
channel-inoculated in the morning (~10 am) with 1 ml of a
Fusarium graminearum macroconidial suspension (500,000
spores/ml) and whole ears were collected and immediately
frozen in liquid nitrogen 6 days later."
BASE COUNT 103 a 102 c 122 g 87 t 22 others
ORIGIN
Query Match 92.0%; Score 18.4; DB 12; Length 436;
Best Local Similarity 95.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGATGAGGACGCGCTGTG 20
|||||
Db 256 CGATGAGGACGCGCTGTG 275

RESULT 9
AW202460/c
LOCUS AW202460
DEFINITION AW202460.1 GI:6501087
ACCESSION AW202460
VERSION AW202460.1
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 444)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 687030 row: G column: 10.
Location/Qualifiers
1. 444
/organism="Zea mays"
/cultivar="Illinois High Oil"
/db_xref="taxon:4577"
/clone_lib="687" - Early embryo from Delaware"
/tissue_type="embryo"
/dev_stage="14, 21, 28, and 35 days after pollination"
/lab_host="E. coli SOLR"
/Note="Organ: embryo; Vector: pBluescript SK; Site_1: XhoI
; Site_2: EcoRI; Library was prepared by Statagene using
the Uni-ZAP XR system (Stratagene BN937328-12). Clones
were picked by a Q-bot after blue/white selection
(ampicillin resistance - use 100 micrograms/microliter).
```

Developed from a pool of equal amounts of RNA from developing embryos sampled at 14, 21, 28 and 35 days after pollination of the Illinois High Oil Maize Strain Cycle 90. This closed strain has been selected for high oil concentration for 90 generations and originates from the 1890s era open pollinated variety Burr's White"

BASE COUNT  
ORIGIN

127 a 116 c 103 g 97 t

Query Match 92.0%; Score 18.4; DB 10; Length 444;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CGATGAGCAGCGCTGTG 20  
|||||  
Db 280 CGATGAGCAGCGCTGTG 261

RESULT 10  
B0635402/c

LOCUS  
DEFINITION B0635402 464 bp mRNA linear EST 15-JUL-2002  
Schmidt lab Zea mays cDNA, mRNA sequence.

ACCESSION B0635402  
VERSION B0635402.1 GI:21759861  
KEYWORDS EST.  
SOURCE Zea mays.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. Walbot, V. 1 (bases 1 to 464)

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu

FEATURES

source  
Location/Qualifiers  
1..464

/organism="Zea mays"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/clone\_lib="1091 - Immature ear with common ESTs screened by Schmidt lab"  
/tissue\_type="Inflorescence meristem - floral organ primordia"  
/dev\_stage="0.5 cm to 2 cm"  
/lab\_host="Stratagene XL0LR"  
/note="Organ: Immature ear; Vector: PAD-GAL4; Site:1: ECORI; Site:2: XhoI; RNA from library 606 was filtered for common ESTs found in 606."

BASE COUNT  
ORIGIN

132 a 119 c 109 g 104 t

Query Match 92.0%; Score 18.4; DB 14; Length 464;  
Best Local Similarity 95.0%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CGATGAGCAGCGCTGTG 20  
|||||  
Db 287 CGATGAGCAGCGCTGTG 268

RESULT 11  
BU037446

LOCUS  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA

BU037446

466 bp mRNA linear

EST 23-AUG-2002

DEFINITION

946139A11.y1 946 - tassell primordium prepared by Schmidt lab Zea mays cDNA, mRNA sequence.  
ACCESSION BU037446  
VERSION BU037446.1 GI:22472966  
KEYWORDS EST.  
SOURCE Zea mays.  
ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. Walbot, V. 1 (bases 1 to 466)

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT  
Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 946139 row: A column: 11.

FEATURES  
source

Location/Qualifiers  
1..466  
/organism="Zea mays"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/clone\_lib="946 - tassell primordium prepared by Schmidt lab"

/tissue\_type="tassels"  
/dev\_stage="just after the transition from vegetative to inflorescence development"  
/lab\_host="XL0LR"

/note="Organ: tassels; Vector: HybrizAP; Site:1: ECORI; Site:2: XhoI; George Chuck dissected immature tassels between imm and 3mm. Sharon Stanfield prepared the cDNA library in HybrizAP. Sample insert size range was 350 bp to 3 Kb with a 1 Kb average."  
BASE COUNT 109 a 115 c 120 g 122 t  
ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 466;  
Best Local Similarity 95.0%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 CGATGAGCAGCGCTGTG 20  
|||||  
Db 216 CGATGAGCAGCGCTGTG 235

RESULT 12  
BI431034/c

LOCUS  
DEFINITION BI431034 481 bp mRNA linear EST 20-AUG-2001  
949064D12.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose Zea mays cDNA, mRNA sequence.

ACCESSION BI431034  
VERSION BI431034.1 GI:15214531  
KEYWORDS EST.  
SOURCE Zea mays.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 481)

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
COMMENT  
Maize ESTs from various cDNA libraries sequenced at Stanford University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA

RESULT 11  
BU037446

LOCUS



BQ294319/c  
 LOCUS BQ294319 498 bp mRNA linear EST 15-MAY-2002  
 DEFINITION 1091028B06.x2 1091 - Immature ear with common ESTs screened by  
 Schmidt lab Zea mays cDNA, mRNA sequence.  
 ACCESSION BQ294319  
 VERSION BQ294319.1 GI:20803269  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE 1 (bases 1 to 498)  
 AUTHORS Walbot, V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford  
 UNIVERSITY  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 1091028 row: B column: 06.  
 FEATURES  
 source  
 1..498  
 /organism="Zea mays"  
 /cultivar="OH43"  
 /db\_xref="taxon:4577"  
 /clone\_lib="1091 - Immature ear with common ESTs screened  
 by Schmidt lab"  
 /tissue\_type="Inflorescence meristem - floral organ  
 primordia"  
 /dev\_stage="0.5 cm to 2 cm"  
 /lab\_host="Stratagene XLOLR"  
 /note="Organ: Immature ear; Vector: pAD-GAL4; Site\_1:  
 EcoRI; Site\_2: XhoI; RNA from library 606 was filtered for  
 common ESTs found in 606."  
 BASE COUNT 130 a 133 c 121 g 114 t  
 ORIGIN

Query Match 92.0%; Score 18.4; DB 14; Length 498;  
 Best Local Similarity 95.0%; Pred. No. 2.2e+02;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CGATGAGGACGCCCGTTG 20  
 ||||||||| |||||||  
 Db 261 CGATGAGGACGCCCGTTG 242

Search completed: July 8, 2003, 09:21:47  
 Job time : 1018.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 125.659 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-14

Perfect score: 20

Sequence: 1 tgcaggaaagggttgctg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

N\_Geneseq\_101002:\*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAZ28820	Rat membrane metal
2	20	100.0	2286	24	Soluble secreted e
3	20	100.0	2765	20	Rat membrane metal
4	18.4	92.0	1529	21	Human Nrbx gene fr
5	18.4	92.0	2061	21	Human nerve differ
6	18.4	92.0	2494	21	Human T-box transc
7	18.4	92.0	2925	21	CDNA encoding neut
8	17.4	87.0	3323	19	Saccharomyces cere
9	16.8	84.0	402	22	Colon tumour relat

10	16.8	84.0	14124	22	ABAI19503
11	16.8	84.0	16854	22	AAS25953
12	16	80.0	656	24	ABK63134
13	16	80.0	1481	22	AAK70269
14	15.8	79.0	200	22	ABA73518
15	15.8	79.0	200	22	ABA38801
16	15.8	79.0	200	22	AAK21963
17	15.8	79.0	200	22	AAK48126
18	15.8	79.0	200	22	AAI53956
19	15.8	79.0	200	22	ABS22034
20	15.8	79.0	358	23	AAS92643
21	15.8	79.0	392	20	AAK40291
22	15.8	79.0	392	21	AAK05013
23	15.8	79.0	397	21	AAA44006
24	15.8	79.0	402	24	AAD30468
25	15.8	79.0	412	23	ABV14145
26	15.8	79.0	434	23	ABV14210
27	15.8	79.0	446	23	ABV35239
28	15.8	79.0	450	23	ABV35304
29	15.8	79.0	450	23	ABV44078
30	15.8	79.0	450	23	ABV44138
31	15.8	79.0	523	21	AAC36791
32	15.8	79.0	525	19	AAV09111
33	15.8	79.0	525	20	AAK77542
34	15.8	79.0	525	21	AAZ59484
35	15.8	79.0	551	22	ABA61095
36	15.8	79.0	551	22	ABA28992
37	15.8	79.0	551	22	AAK09388
38	15.8	79.0	551	22	AAK35277
39	15.8	79.0	551	22	AAI40994
40	15.8	79.0	551	24	ABS09687
41	15.8	79.0	762	24	AAD30459
42	15.8	79.0	849	24	AAD24133
43	15.8	79.0	1464	24	ABK72741
44	15.8	79.0	1467	24	AAD30461
45	15.8	79.0	1543	22	AAH13780

#### ALIGNMENTS

#### RESULT 1

AAZ28820  
ID AAZ28820 standard; DNA; 20 BP.

XX AAZ28820;

AC AAZ28820;

XX 01-FEB-2000 (first entry)

DE Rat membrane metalloprotease NEPII gene probe #10.

XX Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
cardiovascular disease; neurodegenerative disease; growth disorder;  
hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.

XX Synthetic.

OS Rattus rattus.

XX FR2777291-A1.

XX 15-OCT-1999.

XX 08-APR-1998; 98FR-0004389.

XX 08-APR-1998; 98FR-0004389.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;

XX WPI; 1999-593429/51.

DR

XX New membrane metalloprotease NEP II, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease  
 XX  
 PS Claim 3; Page 22; 29pp; French.  
 XX  
 CC Sequences AAZ28811-428827 represent probes for detecting the rat  
 CC membrane metalloprotease designated neprilysine II (NEPII) gene  
 CC (AAZ28810). NEPII is involved in (in)activation of neuronal and hormonal  
 CC peptide messengers. NEPII is used to screen for specific substrates (used  
 CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
 CC used to detect NEPII or for treatment of disorders related to peptidergic  
 CC signalling in which NEPII is involved, e.g. cardiovascular or  
 CC neurodegenerative diseases; growth disorders of endocrine origin;  
 CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
 CC disorders.  
 XX  
 SQ Sequence 20 BP; 4 A; 2 C; 8 G; 6 T; 0 other;  
 Query Match 100.0%; Score 20; DB 20; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGCAGGAAAGGTTGGTCTG 20  
 Db 1 TGCAGGAAAGGTTGGTCTG 20  
 RESULT 2  
 AAZ28130  
 ID AAZ28130 standard; DNA; 2286 BP.  
 XX  
 AC AAZ28130;  
 XX  
 DT 07-MAY-2002 (first entry)  
 XX  
 DE Soluble secreted endopeptidase (SEP) consensus DNA  
 XX  
 KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
 KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
 KW FSAD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
 XX  
 OS Homo sapiens.  
 OS Mus sp.  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 1664..2286  
 FT /\*tag= a  
 FT /note= "Encodes catalytic domain".  
 FT  
 XX WO200206492-A1.  
 XX  
 XX 24-JAN-2002.  
 XX  
 XX 16-JUL-2001; 2001WO-IB01263.  
 XX  
 XX 14-JUL-2000; 2000GB-0017387.  
 XX  
 XX (PFIZ ) PFIZER LTD.  
 XX (PFIZ ) PFIZER INC.  
 XX  
 XX Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
 XX  
 XX WPI; 2002-155042/20.  
 XX  
 XX An isolated and/or purified nucleic acid encoding a human soluble  
 PT secreted endopeptidase which is useful for treating sexual dysfunction,  
 PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
 PT female sexual arousal disorder -  
 XX

PS Disclosure; Fig 6; 167pp; English.  
 XX  
 CC The invention relates to an isolated and/or purified nucleic acid  
 CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
 CC antibody and the compound which inhibits or selectively inhibits the  
 CC human SEP protein are useful in the manufacture of a medicament for  
 CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
 CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
 CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
 CC also useful for treating the above disorders and other disorders such  
 CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
 CC hypoactive sexual desire disorder. The present sequence is SEP consensus  
 CC DNA sequence found in human, mouse and rat.  
 XX  
 SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
 Query Match 100.0%; Score 20; DB 24; Length 2286;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGCAGGAAAGGTTGGTCTG 20  
 Db 982 TGCAGGAAAGGTTGGTCTG 1001  
 RESULT 3  
 AAZ28810  
 ID AAZ28810 standard; cDNA; 2765 BP.  
 XX  
 AC AAZ28810;  
 XX  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Rat membrane metalloprotease NEPII gene.  
 XX  
 KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
 KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
 KW cardiovascular disease; neurodegenerative disease; growth disorder;  
 KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
 XX  
 OS Rattus rattus.  
 XX  
 PN FR2777291-A1.  
 XX  
 PD 15-OCT-1999.  
 XX  
 PF 08-APR-1998; 98FR-0004389.  
 XX  
 PR 08-APR-1998; 98FR-0004389.  
 XX  
 PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
 PI Schwartz JC;  
 XX  
 DR WPI; 1999-593429/51.  
 DR P-PSDB; AAY44177.  
 XX  
 XX New membrane metalloprotease NEPII, involved in proteolysis of  
 PT neuronal and hormonal peptides, used to screen for inhibitors,  
 PT potentially useful for treating e.g. cardiovascular disease -  
 XX  
 XX Claim 2; Page 12-16; 29pp; French.  
 XX  
 CC This sequence represents the gene for the rat membrane metalloprotease  
 CC designated neprilysine II (NEPII), which is involved in (in)activation  
 CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
 CC for specific substrates (used to detect NEPII in cells and tissues) or  
 CC inhibitors, which can also be used to detect NEPII or for treatment of  
 CC disorders related to peptidergic signalling in which NEPII is involved,  
 CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
 CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
 CC endocrine disorders.

```

XX SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;
Query Match 100.0%; Score 20; DB 20; Length 2765;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 1110 TGCAGGAAAGGTTGGTCTG 1129

RESULT 4
AAZ46480/c
ID AAZ46480 standard; DNA; 1529 BP.
XX AC AAZ46480;
XX DT 13-MAR-2000 (first entry)
XX DE Human MTbx gene fragment.
XX KW T-box transcription factor protein; MTbx; immune system disease; HIV;
XX KW leukemia; inflammatory disease; asthma; rheumatoid arthritis; human;
XX KW developmental disorder; cardiovascular; transcriptional regulation;
XX KW myocardial ischemia; gene mapping; ss.
XX OS Homo sapiens.
XX PN WO9966025-A2.
XX PD 23-DEC-1999.
XX PF 15-JUN-1999; 99WO-US13457.
XX PR 16-JUN-1998; 98US-0089467.
XX PR 29-SEP-1998; 98US-0163116.
XX PR 09-NOV-1998; 98US-0188811.
XX PR 10-NOV-1998; 98US-0189760.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Rhodadoust M;
XX DR WPI; 2000-097738/08.
XX DR P-PSDB; AAY49415.
XX PT Nucleotides encoding human T-box transcription factor protein, MTbx,
XX PS useful to treat transcriptional regulation disorders.
XX PS Disclosure; Page 148-150; 152pp; English.
XX CC The invention relates to a human T-box transcription factor protein,
XX CC MTbx. The polypeptide can be administered therapeutically, especially by
XX CC expressing encoding polynucleotides, to treat disorders characterized by
XX CC aberrant polypeptide expression or activity, e.g. immune system diseases
XX CC (e.g. HIV, leukemia, etc.); chronic inflammatory diseases (e.g. asthma,
XX CC rheumatoid arthritis etc.); developmental disorders, cardiovascular
XX CC disorders, or other disorders arising from improper transcriptional
XX CC regulation. They are especially useful to treat conditions characterized
XX CC by loss of tissue integrity relating to disease and/or injury e.g. HIV,
XX CC hibernating myocardium during myocardial ischemia etc. It can be used to
XX CC diagnose diseases related to polypeptide expression or activity. The
XX CC polynucleotide may also be used to produce non-human transgenic animals
XX CC and to produce probes useful to detect related sequences, or for gene
XX CC mapping. The present sequence represents a fragment of the human MTbx
XX CC gene.
XX SQ Sequence 1529 BP; 432 A; 342 C; 335 G; 416 T; 4 other;
Query Match 92.0%; Score 18.4; DB 21; Length 1529;
Best Local Similarity 95.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

XX SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;
Query Match 100.0%; Score 20; DB 20; Length 2765;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 1110 TGCAGGAAAGGTTGGTCTG 1129

RESULT 5
AAZ90785/c
ID AAZ90785 standard; DNA; 2061 BP.
XX AC AAZ90785;
XX DT 23-JUN-2000 (first entry)
XX DE Human nerve differentiation factor, Tbr-2 encoding DNA.
XX KW Tbr-2; nerve differentiation factor; brain; nerve disorder; human;
XX KW senile dementia; nerve regeneration; nerve injury; ds.
XX OS Homo sapiens.
XX PH Key Location/Qualifiers
XX FT CDS 1..2061
XX FT /*tag= a
XX FT /product= "Tbr-2"
XX PN WO200012712-A1.
XX PD 09-MAR-2000.
XX PF 18-JUN-1999; 99WO-JP03271.
XX PR 26-AUG-1998; 98JP-0240920.
XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Kimura N, Taga T;
XX DR WPI; 2000-237874/20.
XX DR P-PSDB; AAY57394.
XX PT Gene Tbr-2 highly expressed in fetal brain tissue encodes a nerve
XX PT differentiation factor for treatment of nerve disorders such as
XX PT dementia and nerve injury.
XX PS Claim 2; Page 75-82; 89pp; Japanese.
XX CC The invention relates to a gene Tbr-2 encoding a nerve differentiation
XX CC factor that is highly expressed in fetal brain tissue. The nerve
XX CC differentiation factor polypeptides can be expressed by standard
XX CC recombinant methodology. The nerve differentiation factor is useful for
XX CC the treatment of nerve disorders such as senile dementia, and for the
XX CC promotion of nerve regeneration after nerve injury. The present sequence
XX CC represents a human Tbr-2 gene encoding the nerve differentiation factor.
XX SQ Sequence 2061 BP; 465 A; 640 C; 557 G; 399 T; 0 other;
Query Match 92.0%; Score 18.4; DB 21; Length 2061;
Best Local Similarity 95.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 1713 TGCAGGAAAGGTTGGTCTG 1694

RESULT 6
AAZ46477/c
ID AAZ46477 standard; cDNA; 2494 BP.
XX AC AAZ46477;
XX DT 13-MAR-2000 (first entry)

```

XX DE Human T-box transcription factor protein, MTBx encoding cDNA.  
 XX KW T-box transcription factor protein; MTBx; immune system disease; HIV;  
 KW leukemia; inflammatory disease; asthma; rheumatoid arthritis; human;  
 KW developmental disorder; cardiovascular; transcriptional regulation;  
 KW myocardial ischemia; gene mapping; chromosome 3p23-p24; ss.  
 XX OS Homo sapiens.  
 XX FH Key Location/Qualifiers  
 FT CDS 164..1717  
 FT /\*tag= a  
 FT /product= "MTbx"  
 XX WO9966025-A2.  
 XX PD 23-DEC-1999.  
 XX PF 15-JUN-1999; 99WO-US13457.  
 XX PR 16-JUN-1998; 98US-0089467.  
 PR 29-SEP-1998; 98US-0163116.  
 PR 09-NOV-1998; 98US-0188811.  
 PR 10-NOV-1998; 98US-0189760.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX PA Rhodadoust M;  
 XX PI WPI: 2000-097738/08.  
 DR P-PSDB; AA49414.  
 XX PT Nucleotides encoding human T-box transcription factor protein, MTbx,  
 PT useful to treat transcriptional regulation disorders  
 XX Claim 2; Fig 1; 152pp; English.  
 XX CC This cDNA encodes a human T-box transcription factor protein, MTbx. The  
 CC polypeptide can be administered therapeutically, especially by expressing  
 CC encoding polynucleotides, to treat disorders characterized by aberrant  
 CC polypeptide expression or activity, e.g. immune system diseases (e.g.  
 CC HIV, leukemia, etc.), chronic inflammatory diseases (e.g. asthma,  
 CC rheumatoid arthritis etc.), developmental disorders, cardiovascular  
 CC disorders, or other disorders arising from improper transcriptional  
 CC regulation. They are especially useful to treat conditions characterized  
 CC by loss of tissue integrity relating to disease and/or injury e.g. HIV,  
 CC hibernating myocardium during myocardial ischemia etc. It can be used to  
 CC diagnose diseases related to polypeptide expression or activity. The  
 CC polynucleotide may also be used to produce non-human transgenic animals  
 CC and to produce probes useful to detect related sequences, or for gene  
 CC mapping. The MTbx gene was mapped to chromosome 3p23-p24.  
 XX SQ Sequence 2494 BP; 651 A; 639 C; 600 G; 594 T; 10 other;  
 Query Match 92.0%; Score 18.4; DB 21; Length 2494;  
 Best Local Similarity 95.0%; Pred. No. 15;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TGCAGGAAAGGTTGGTCTG 20  
 |||||  
 Db 1369 TGCAGGAAAGGTTGGTCTG 1350  
 RESULT 7  
 ID AAA63763  
 XX AAA63763 standard; cDNA; 2925 BP.  
 AC AAA63763;  
 XX 04-DEC-2000 (first entry)  
 DE CDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.

XX NEprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
 KW NEP-like enzyme; protein production; protein secretion;  
 KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
 KW fertility; bone disease; abnormal phosphate metabolism; ss.  
 XX OS Mus sp.  
 XX FH Key Location/Qualifiers  
 FT CDS 332..2629  
 FT /\*tag= a  
 FT /product= "neutral endopeptidase metalloproteinase-like  
 enzyme NL-1"  
 XX WO200047750-A2.  
 XX PD 17-AUG-2000.  
 XX PF 11-FEB-2000; 2000WO-CA00147.  
 XX PR 11-FEB-1999; 99CA-2260376.  
 XX (UYMO-) UNIV MONTREAL.  
 XX PI Desgroseillers L, Boileau G;  
 XX WPI: 2000-549148/50.  
 DR P-PSDB; AAB08130.  
 XX PT Novel neutral endopeptidase-like metalloproteinase polypeptides and  
 PT polynucleotides, used to screen for related sequences and enzyme  
 PT inhibitors, used for the treatment of NL-3 related bone disorders  
 XX Disclosure; Fig 3; 59pp; English.  
 XX CC The present sequence encodes a murine neutral endopeptidase  
 CC metalloproteinase-like enzyme, designated NL-1. The specification  
 CC also describes NL-2 and NL-3. The NL enzymes are used to test for  
 CC specific inhibitors. The N-terminal region of the enzymes can be used  
 CC to promote production and secretion of foreign proteins and active  
 CC polypeptides, using chimeric constructs containing the foreign protein  
 CC downstream from and in phase with the N-terminal region. The NL enzymes  
 CC are have been localised to the brain, and may be useful in the  
 CC treatment of neurological diseases such as Alzheimer's disease, pain,  
 CC and psychiatric disorders. NL enzymes have also been localised to the  
 CC testis and ovaries, and may be used to control fertility. They have  
 CC also been localised to bones, and may be used to treat bone diseases,  
 CC and abnormal phosphate metabolisms related to improper peptide  
 CC processing by the NL-3 enzyme.  
 XX SQ Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;  
 Query Match 92.0%; Score 18.4; DB 21; Length 2925;  
 Best Local Similarity 95.0%; Pred. No. 16;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TGCAGGAAAGGTTGGTCTG 20  
 |||||  
 Db 1308 TACAGGAAAGGTTGGTCTG 1327  
 RESULT 8  
 ID AAV40294/c  
 XX AAV40294 standard; DNA; 3323 BP.  
 AC AAV40294;  
 XX 15-OCT-1998 (first entry)  
 DE Saccharomyces cerevisiae LIG4 encoding nucleotide sequence.  
 KW DNA ligase IV; LIG4; cellular DNA repair; modulation; diagnosis;  
 KW XRC4; Ku-associated DNA repair pathway; cancer; retroviral infection;



ID	ABR19503 standard; DNA; 14124 BP.	08-SEP-2000; 2000US-0232080.
AC	ABR19503;	PR 08-SEP-2000; 2000US-0232081.
AC	ABR19503;	PR 12-SEP-2000; 2000US-0231968.
XX		PR 14-SEP-2000; 2000US-0232397.
DT	23-JAN-2002 (first entry)	PR 14-SEP-2000; 2000US-0232398.
XX		PR 14-SEP-2000; 2000US-0232399.
XX		PR 14-SEP-2000; 2000US-0232400.
DE	Human nervous system related polynucleotide SEQ ID NO 11834.	PR 14-SEP-2000; 2000US-0232401.
XX		PR 14-SEP-2000; 2000US-0233063.
KW	Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;	PR 14-SEP-2000; 2000US-0233064.
KW	immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;	PR 14-SEP-2000; 2000US-0233065.
KW	antiparkinsonian; antispasmodic; antianemic; antithrombotic; cancer;	PR 21-SEP-2000; 2000US-0234223.
KW	antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;	PR 21-SEP-2000; 2000US-0234274.
KW	antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;	PR 25-SEP-2000; 2000US-0234997.
KW	antiparasitic; cardiant; immune disorder; cardiovascular disorder;	PR 25-SEP-2000; 2000US-0234998.
KW	neurological disease; infection; gene therapy; vaccine; ds.	PR 26-SEP-2000; 2000US-0235484.
XX		PR 27-SEP-2000; 2000US-0235834.
OS	Homo sapiens.	PR 27-SEP-2000; 2000US-0235836.
XX		PR 29-SEP-2000; 2000US-0236327.
PN	WO200159063-A2.	PR 29-SEP-2000; 2000US-0236367.
XX		PR 29-SEP-2000; 2000US-0236368.
XX		PR 29-SEP-2000; 2000US-0236369.
PF	16-AUG-2001.	PR 29-SEP-2000; 2000US-0236370.
XX		PR 02-OCT-2000; 2000US-0236802.
XX	17-JAN-2001; 2001WO-US01334.	PR 02-OCT-2000; 2000US-0237037.
XX		PR 02-OCT-2000; 2000US-0237038.
PR	31-JAN-2000; 2000US-0179065.	PR 02-OCT-2000; 2000US-0237039.
PR	04-FEB-2000; 2000US-0180528.	PR 12-OCT-2000; 2000US-0237040.
PR	24-FEB-2000; 2000US-0184664.	PR 13-OCT-2000; 2000US-0239935.
PR	02-MAR-2000; 2000US-0186350.	PR 13-OCT-2000; 2000US-0239937.
PR	16-MAR-2000; 2000US-0189874.	PR 13-OCT-2000; 2000US-0240960.
PR	17-MAR-2000; 2000US-0190076.	PR 20-OCT-2000; 2000US-0241785.
PR	18-APR-2000; 2000US-0198123.	PR 20-OCT-2000; 2000US-0241786.
PR	19-MAY-2000; 2000US-0205515.	PR 20-OCT-2000; 2000US-0241787.
PR	07-JUN-2000; 2000US-0209467.	PR 20-OCT-2000; 2000US-0241808.
PR	28-JUN-2000; 2000US-0214986.	PR 20-OCT-2000; 2000US-0241809.
PR	30-JUN-2000; 2000US-0215135.	PR 20-OCT-2000; 2000US-0241826.
PR	07-JUL-2000; 2000US-0216847.	PR 01-NOV-2000; 2000US-0244617.
PR	07-JUL-2000; 2000US-0216880.	PR 08-NOV-2000; 2000US-0246474.
PR	11-JUL-2000; 2000US-0217487.	PR 08-NOV-2000; 2000US-0246475.
PR	11-JUL-2000; 2000US-0217496.	PR 08-NOV-2000; 2000US-0246476.
PR	14-JUL-2000; 2000US-0218290.	PR 08-NOV-2000; 2000US-0246477.
PR	26-JUL-2000; 2000US-0220963.	PR 08-NOV-2000; 2000US-0246478.
PR	26-JUL-2000; 2000US-0220964.	PR 08-NOV-2000; 2000US-0246521.
PR	14-AUG-2000; 2000US-0224518.	PR 08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0224519.	PR 08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225213.	PR 08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225214.	PR 08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225266.	PR 08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225267.	PR 08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225268.	PR 08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225270.	PR 08-NOV-2000; 2000US-0246610.
PR	14-AUG-2000; 2000US-0225447.	PR 08-NOV-2000; 2000US-0246611.
PR	14-AUG-2000; 2000US-0225757.	PR 08-NOV-2000; 2000US-0246613.
PR	14-AUG-2000; 2000US-0225758.	PR 17-NOV-2000; 2000US-0249207.
PR	14-AUG-2000; 2000US-0225759.	PR 17-NOV-2000; 2000US-0249208.
PR	18-AUG-2000; 2000US-0226279.	PR 17-NOV-2000; 2000US-0249209.
PR	22-AUG-2000; 2000US-0226681.	PR 17-NOV-2000; 2000US-0249210.
PR	22-AUG-2000; 2000US-0226682.	PR 17-NOV-2000; 2000US-0249211.
PR	22-AUG-2000; 2000US-0226686.	PR 17-NOV-2000; 2000US-0249212.
PR	22-AUG-2000; 2000US-0227182.	PR 17-NOV-2000; 2000US-0249214.
PR	23-AUG-2000; 2000US-0227009.	PR 17-NOV-2000; 2000US-0249215.
PR	30-AUG-2000; 2000US-0228924.	PR 17-NOV-2000; 2000



PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-)	HUMAN GENOME SCI INCI	
PA			
XX	Rosen CA, Barash SC, Ruben		
PI	WPI: 2001-457723/49.		
XX			
DR	Isolated polypeptide for treat		
XX	respiratory disorders relate		
PT	also for testing and detecti		
PT			
XX	Claim 1; SEQ ID: No 217; 507p		
PS	Sequences AAS29931-AAS30164		
CC	the lung antigen polypeptid		
CC	and their associated polynuc		
CC	treatment and prevention o		
CC	mice, rabbits, goats, horse		
CC	pathological condition can b		
CC	absence of a mutation in a l		
CC	disorders include autoimmun		
CC	hyperproliferative disorder		
CC	cardiovascular disorders suc		
CC	disorders such as cerebral i		
CC	Alzheimer's disease, infecti		
CC	ocular disorders such as cor		
CC	premature labour and infert		
CC	Crohn's disease, renal disor		
CC	respiratory disorders, such a		
CC	also be used to aid wound he		
CC	to maintain organs before tr		
CC	chemotaxis. The polypeptid		
CC	preservative to increase o		
CC	Note: The sequence data for		
CC	specification, but was obta		
CC	at ftp.wipo.int/pub/publish		
CC			
Query Match	84.0%;		
Best Local Similarity	90.0%;		
Matches	18; Conservative		
QY	1 TGCAGGAAGAGTTTGCTCTC		
Db	5587 TGCAGGACAGTTTGTTT		
RESULT 12			
ID	ABK63134		
XX	AC		
XX	ABK63134;		
XX	AC		
DT	18-JUN-2002. (first entry)		
DE	Rat sequence differentially		
XX			
KW	Rat; ss; hepatotoxin; expres		
KW	differential expression; cer		
XX			
OS	Rattus norvegicus.		
XX			
PN	WO200210453-A2.		
PD	07-FEB-2002.		
XX			
PF	30-JUL-2001; 2001WO-US23872		
XX			
XX	31-JUL-2000; 2000US-222040P		
PR	02-NOV-2000; 2000US-244880P		

QY 1 TGCAGGAAGGTTTGGTCTG 20  
||||| } ||||| } ||  
Db 5587 TGCAGGACAGGTTTGGTTTG 5568

RESULT 12  
ABK63134  
ID ABK63134 standard; cDNA; 656 BP.  
XX AC ABK63134;  
XX DT 18-JUN-2002. (first entry)  
XX DE Rat sequence differentially expressed in response to a hepatotoxin #1041.  
XX KW Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
XX KW differential expression; centrilobular necrosis; steatosis.  
XX OS Rattus norvegicus.  
XX PN WO200210453-A2.  
XX PD 07-FEB-2002.  
XX PF 30-JUL-2001; 2001WO-US23872.  
XX PR 31-JUL-2000; 2000US-222040P.  
PR 02-NOV-2000; 2000US-244880P.

PR 11-MAY-2001; 2001US-290029P.  
 PR 15-MAY-2001; 2001US-290645P.  
 PR 22-MAY-2001; 2001US-292336P.  
 PR 06-JUN-2001; 2001US-295798P.  
 PR 13-JUN-2001; 2001US-297457P.  
 PR 19-JUN-2001; 2001US-298884P.  
 PR 09-JUL-2001; 2001US-303459P.  
 XX XX

(GENE-) GENE LOGIC INC.

PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;

XX WPI; 2002-241625/29.

XX Predicting toxic effects of compounds or the progression of these toxic  
 PT effects by determining the changes in gene expression in tissues or  
 PT cells exposed to the toxin and comparing these to gene expression in  
 PT unexposed tissues or cells

PS Claim 1; Seq ID No 1041; 239pp; English.

XX The invention relates to methods for predicting toxic effects of  
 CC compounds or the progression of these toxic effects by determining the  
 CC global changes in gene expression in tissues or cells exposed to the  
 CC toxin and comparing these to gene expression in unexposed tissues or  
 CC cells. Also included are methods of predicting at least one toxic  
 CC effect of a compound or progression of a toxic effect, preferably the  
 CC hepatotoxicity of a compound, comprising detecting the level of  
 CC expression in a tissue or cell sample exposed to the compound of two or  
 CC more genes listed in the specification, where differential expression of  
 CC the genes is indicative of at least one toxic effect or progression.  
 CC The method can also be used to identify an agent which modulates the  
 CC toxic response and predict cellular pathways that a compound modulates  
 CC in a cell. The methods utilize a set of at least two probes (on a solid  
 CC support in kit form), where each of the probes comprises a sequence that  
 CC specifically hybridises to a gene listed in the specification, a computer  
 CC system comprising a database containing information identifying the  
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
 CC set of genes comprising at least two genes listed in the specification,  
 CC and a user interface to view the information used to present information,  
 CC identifying the expression level in a tissue or cell of at least one gene  
 CC listed in the specification. The method is useful for elucidating global  
 CC changes in gene expression and for identifying toxicity markers in  
 CC tissues or cell exposed to a known toxin. The genes may be used as  
 CC toxicity markers in drug screening and toxicity assays. The genes and  
 CC gene expression information may be used as diagnostic markers for the  
 CC prediction or identification of the physiological state of tissue or cell  
 CC sample that has been exposed to a compound or agent. Hepatotoxicity  
 CC is characterised by centrilobular necrosis and steatosis. The present  
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
 CC which is differentially expressed in response to a hepatotoxic agent.

SQ Sequence 656 BP; 151 A; 120 C; 206 G; 179 T; 0 other;

Query Match 80.0%; Score 16; DB 24; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AGGAAAGGTTGGTCT 19  
 |||||  
 Db 596 AGCAAGGTTGGTCT 611

RESULT 13

AAK70269/C

ID AAK70269 standard; DNA: 1481 BP.

XX AAK70269;

XX 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25081.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.  
 OS Homo sapiens.  
 PN WO200157182-A2.  
 XX 09-AUG-2001.  
 PD 17-JAN-2001; 2001WO-US01354.  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.

```
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis
XX
XX PS Disclosure; SEQ ID NO 25081; 307lpp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX SQ Sequence 1481 BP; 404 A; 334 C; 316 G; 427 T; 0 other;
Query Match 80.0%; Score 16; DB 22; Length 1481;
Best Local Similarity 100.0%; Pred. No. 2.3e+02; Mismatches 0; Gaps 0;
Matches 16; Conservative 0; Indels 0;

QY 5 GGAAAGGTTTGGTCTG 20
IIIIIIIIIIIIIIII
Db 1205 GGAAAGGTTTGGTCTG 1190

RESULT 14
ABA73518/c
ID ABA73518 standard; DNA; 200 BP.
XX
XX AC ABA73518;
XX
XX DT 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #21823.
XX
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX
XX PN WO200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00669.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
```

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver  
XX  
XX Claim 4; SEQ ID NO 21823; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human fetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 200 BP; 40 A; 54 C; 44 G; 62 T; 0 other;  
SQ  
Query Match 79.0%; Score 15.8; DB 22; Length 200;  
Best Local Similarity 89.5%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCAGGAAAGGTTTGGTCT 19  
Db 129 TGCAGGAAAGGTTTGGCCT 111

RESULT 15  
ABA3801/C  
ID ABA3801 standard; DNA; 200 BP.  
XX  
XX ABA3801;  
XX  
XX 23-JAN-2002 (first entry)  
XX  
XX Probe #17267 for gene expression analysis in human heart cell sample.  
XX Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157274-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00666.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
XX Claim 4; SEQ ID No 17267; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The

CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 200 BP; 40 A; 54 C; 44 G; 62 T; 0 other;  
SQ  
Query Match 79.0%; Score 15.8; DB 22; Length 200;  
Best Local Similarity 89.5%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGCAGGAAAGGTTTGGTCT 19  
Db 129 TGCAGGAAAGGTTTGGCCT 111

Search completed: July 8, 2003, 02:19:02  
Job time : 127.659 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 01:24:03 ; Search time 27.6098 Seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-14

Perfect score: 20

Sequence: 1 tgcaggaagggttgctg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Issued\_Patents\_NA: \*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2.6/ptodata/1/ina/PTUS\_COMB.seq: \*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
C 1	18.4	92.0	1529	3	US-09-189-760-5		Sequence 5, Appl
C 2	18.4	92.0	1529	3	US-09-188-811-5		Sequence 5, Appl
C 3	18.4	92.0	1529	4	US-09-514-422-5		Sequence 5, Appl
C 4	18.4	92.0	1704	3	US-09-188-811-1		Sequence 1, Appl
C 5	18.4	92.0	2494	3	US-09-189-760-1		Sequence 1, Appl
C 6	18.4	92.0	2494	4	US-09-514-422-1		Sequence 1, Appl
C 7	15.8	79.0	525	1	US-08-686-878A-29		Sequence 29, Appl
C 8	15.8	79.0	525	4	US-09-175-928-29		Sequence 29, Appl
C 9	15.4	77.0	466	4	US-09-370-838-217		Sequence 217, Appl
C 10	15.4	77.0	2052	3	US-08-630-916A-45		Sequence 45, Appl
C 11	15.4	77.0	2052	4	US-09-657-481A-3		Sequence 3, Appl
C 12	15.2	76.0	341	3	US-08-545-809A-32		Sequence 32, Appl
C 13	15.2	76.0	554	3	US-08-840-146-16		Sequence 16, Appl
C 14	15.2	76.0	554	3	US-09-360-220-16		Sequence 16, Appl
C 15	15.2	76.0	1237	4	US-08-961-527-325		Sequence 325, Appl
C 16	15.2	76.0	1830	1	US-07-691-079C-3		Sequence 3, Appl
C 17	15.2	76.0	1830	1	US-08-433-037-3		Sequence 3, Appl
C 18	15.2	76.0	2241	3	US-08-840-146-1		Sequence 1, Appl
C 19	15.2	76.0	2241	3	US-09-360-220-1		Sequence 1, Appl
C 20	15.2	76.0	3475	4	US-09-657-481A-10		Sequence 10, Appl
C 21	15.2	76.0	3476	3	US-08-630-916A-47		Sequence 47, Appl
C 22	15.2	76.0	3580	4	US-09-081-345-1		Sequence 1, Appl
C 23	15.2	76.0	21126	1	US-08-008-216-19		Sequence 19, Appl
C 24	15.2	76.0	21126	1	US-08-459-569-19		Sequence 19, Appl
C 25	15.2	76.0	21126	1	US-08-458-831-19		Sequence 19, Appl
C 26	15.2	76.0	43360	4	US-09-453-702B-206		Sequence 206, Appl
C 27	15.2	76.0	45325	4	US-09-453-702B-261		Sequence 261, Appl

28 14.8 74.0 569 4 US-09-328-111-561  
c 29 14.8 74.0 1320 4 US-09-457-046B-44  
c 30 14.8 74.0 1866 1 US-08-336-408B-1  
c 31 14.8 74.0 1866 5 PCT-US91-00399-1  
c 32 14.8 74.0 2095 1 US-08-336-408B-3  
c 33 14.8 74.0 2095 5 PCT-US91-00399-3  
c 34 14.8 74.0 13104 4 US-08-961-527-34  
c 35 14.4 72.0 433 4 US-09-397-787-185  
c 36 14.4 72.0 1757 1 US-08-313-075A-49  
c 37 14.4 72.0 1942 4 US-09-326-203A-1  
c 38 14.4 72.0 5502 3 US-08-836-134-1  
c 39 14.4 72.0 5502 4 US-09-493-784-1  
c 40 14.2 71.0 254 4 US-08-940-424-8  
c 41 14.2 71.0 527 4 US-09-325-932A-44  
c 42 14.2 71.0 851 4 US-08-998-416-323  
c 43 14.2 71.0 931 3 US-09-049-672A-19  
c 44 14.2 71.0 1334 4 US-08-940-424-13  
c 45 14.2 71.0 1578 3 US-08-852-936C-2

#### ALIGNMENTS

##### RESULT 1

US-09-189-760-5/c  
; Sequence 5, Application US/09189760  
; Patent No. 6031078  
; GENERAL INFORMATION:  
; APPLICANT: Khodadoust, Mehran  
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: MNI-046CP2  
; CURRENT APPLICATION NUMBER: US/09/189,760  
; CURRENT FILING DATE: 1998-11-10  
; EARLIER APPLICATION NUMBER: 09/163,116  
; EARLIER FILING DATE: 1998-09-29  
; EARLIER APPLICATION NUMBER: 60/089,467  
; EARLIER FILING DATE: 1998-06-16  
; EARLIER APPLICATION NUMBER: (PENDING)  
; EARLIER FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1529  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (3)..(749)  
US-09-189-760-5

Query Match 92.0%; Score 18.4; DB 3; Length 1529;  
Best Local Similarity 95.0%; Pred. No. 0.89;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGGTTTGCTGTG 20  
|||||  
Db 404 TGCAGGAAAGGTTTGCTGTG 385

##### RESULT 2

US-09-188-811-5/c  
; Sequence 5, Application US/09188811  
; Patent No. 6037148  
; GENERAL INFORMATION:  
; APPLICANT: Khodadoust, Mehran  
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES  
; TITLE OF INVENTION: THEREFOR  
; FILE REFERENCE: MNI-046CP  
; CURRENT APPLICATION NUMBER: US/09/188,811  
; CURRENT FILING DATE: 1998-11-09  
; EARLIER APPLICATION NUMBER: 09/163,116  
; EARLIER FILING DATE: 1998-09-29

```

; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(749)
US-09-188-811-5

Query Match          92.0%; Score 18.4; DB 3; Length 1529;
Best Local Similarity 95.0%; Pred. No. 0.89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 404 TGCAGGAAAGGTTGGTCTG 385

RESULT 3
US-09-514-422-5/c
; Sequence 5, Application US/09514422
; Patent No. 6291193
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MNI-046CP2
; CURRENT APPLICATION NUMBER: US/09/514,422
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US/09/189,760
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 09/163,116
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/089,467
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: (PENDING)
; PRIOR FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1529
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(749)
US-09-514-422-5

Query Match          92.0%; Score 18.4; DB 4; Length 1529;
Best Local Similarity 95.0%; Pred. No. 0.89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 404 TGCAGGAAAGGTTGGTCTG 385

RESULT 4
US-09-188-811-1/c
; Sequence 1, Application US/09188811
; Patent No. 6037148
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MNI-046CP
; CURRENT APPLICATION NUMBER: US/09/188,811
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: 09/163,116
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1704
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1701)
US-09-188-811-1

Query Match          92.0%; Score 18.4; DB 3; Length 1704;
Best Local Similarity 95.0%; Pred. No. 0.91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 1356 TGCAGGAAAGGTTGGTCTG 1337

RESULT 5
US-09-189-760-1/c
; Sequence 1, Application US/09189760
; Patent No. 6031078
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MNI-046CP2
; CURRENT APPLICATION NUMBER: US/09/189,760
; CURRENT FILING DATE: 1998-11-10
; EARLIER APPLICATION NUMBER: 09/163,116
; EARLIER FILING DATE: 1998-09-29
; EARLIER APPLICATION NUMBER: 60/089,467
; EARLIER FILING DATE: 1998-06-16
; EARLIER APPLICATION NUMBER: (PENDING)
; EARLIER FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (164)..(1714)
US-09-189-760-1

Query Match          92.0%; Score 18.4; DB 3; Length 2494;
Best Local Similarity 95.0%; Pred. No. 0.99;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGGTTGGTCTG 20
    |||||
Db 1369 TGCAGGAAAGGTTGGTCTG 1350

RESULT 6
US-09-514-422-1/c
; Sequence 1, Application US/09514422
; Patent No. 6291193
; GENERAL INFORMATION:
; APPLICANT: Khodadoust, Mehran
; TITLE OF INVENTION: NOVEL MTBX PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MNI-046CP2
; CURRENT APPLICATION NUMBER: US/09/514,422
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US/09/189,760
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 09/163,116
; PRIOR FILING DATE: 1998-09-29
; PRIOR APPLICATION NUMBER: 60/089,467
; PRIOR FILING DATE: 1998-06-16
```

;; PRIOR APPLICATION NUMBER: (PENDING)  
;; PRIOR FILING DATE: 1998-11-09  
;; NUMBER OF SEQ ID NOS: 10  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1  
;; LENGTH: 2494  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (164)..(1714)  
US-09-514-422-1

Query Match 92.0%; Score 18.4; DB 4; Length 2494;  
Best Local Similarity 95.0%; Pred. No. 0.99;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAGGTTGGTCTG 20  
Db 1369 TGCAGGAAGGTTGGTCTG 1350

## RESULT 7

US-08-686-878A-29/c  
;; Sequence 29, Application US/08686878A  
;; Patent No. 5708157  
;; GENERAL INFORMATION:

;; APPLICANT: Jacobs, Kenneth  
;; APPLICANT: McCoy, John  
;; APPLICANT: LaVallie, Edward  
;; APPLICANT: Racie, Lisa  
;; APPLICANT: Merberg, David  
;; APPLICANT: Treacy, Maurice  
;; APPLICANT: Evans, Cheryl  
;; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
;; NUMBER OF SEQUENCES: 71  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Genetics Institute, Inc.  
;; STREET: 87 Cambridgepark Drive  
;; CITY: Cambridge  
;; STATE: Massachusetts  
;; COUNTRY: U.S.A.  
;; ZIP: 02140  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/686,878A  
;; FILING DATE:  
;; CLASSIFICATION: 536  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Brown, Scott A.  
;; REGISTRATION NUMBER: 32,724  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 498-8224  
;; TELEFAX: (617) 876-5951  
;; INFORMATION FOR SEQ ID NO: 29:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 525 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
US-08-686-878A-29

Query Match 79.0%; Score 15.8; DB 1; Length 525;  
Best Local Similarity 89.5%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTGGTCTG 20  
Db 388 GCAGGAAGGTTGGGCTG 370

## RESULT 8

US-09-175-928-29/c  
;; Sequence 29, Application US/09175928A  
;; Patent No. 6312921  
;; GENERAL INFORMATION:  
;; APPLICANT: Jacobs, Kenneth  
;; APPLICANT: McCoy, John M.  
;; APPLICANT: LaVallie, Edward R.  
;; APPLICANT: Collins-Racie, Lisa A.  
;; APPLICANT: Evans, Cheryl  
;; APPLICANT: Merberg, David  
;; APPLICANT: Treacy, Maurice  
;; APPLICANT: MI, Sha  
;; APPLICANT: Genetics Institute, Inc.  
;; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
;; FILE REFERENCE: 6006B.AJ172A  
;; CURRENT APPLICATION NUMBER: US/09/175,928A  
;; CURRENT FILING DATE: 1998-10-20  
;; NUMBER OF SEQ ID NOS: 62  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 29  
;; LENGTH: 525  
;; TYPE: DNA  
;; ORGANISM: Mus musculus  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (22)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (52)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (35)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (59)  
US-09-175-928-29

Query Match 79.0%; Score 15.8; DB 4; Length 525;  
Best Local Similarity 89.5%; Pred. No. 17;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTGGTCTG 20  
Db 388 GCAGGAAGGTTGGGCTG 370

## RESULT 9

US-09-370-838-217  
;; Sequence 217, Application US/09370838  
;; Patent No. 6444425  
;; GENERAL INFORMATION:  
;; APPLICANT: Reed, Steven G.  
;; APPLICANT: Lodes, Michael J.  
;; APPLICANT: Mohamath, Roadoh  
;; APPLICANT: Secrist, Heather  
;; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF  
;; FILE REFERENCE: 210121.475C1  
;; CURRENT APPLICATION NUMBER: US/09/370,838  
;; CURRENT FILING DATE: 1999-08-09  
;; EARLIER APPLICATION NUMBER: US 09/285,323  
;; EARLIER FILING DATE: 1999-04-02  
;; NUMBER OF SEQ ID NOS: 289  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 217  
;; LENGTH: 466  
;; TYPE: DNA

```
; ORGANISM: Homo sapien
US-09-370-838-217

Query Match      77.0%; Score 15.4; DB 4; Length 466;
Best Local Similarity 94.1%; Pred. No. 27;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGT 17
    ||||| |||||
Db 343 TGCAGGAAGGTTGGT 359

RESULT 10
US-08-630-916A-45/c
; Sequence 45, Application US/08630916A
; Patent No. 6011137
; GENERAL INFORMATION:
; APPLICANT: Pirozzi, Gregorio
; APPLICANT: Kay, Brian K.
; APPLICANT: Fowlkes, Dana M.
; TITLE OF INVENTION: IDENTIFICATION AND ISOLATION OF NOVEL
; POLYPEPTIDES HAVING WW DOMAINS AND METHODS OF USING SAME.
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,916A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. LESLIE
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-203
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 896-8864/9741
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2052 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-630-916A-45

Query Match      77.0%; Score 15.4; DB 3; Length 2052;
Best Local Similarity 94.1%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AGGAAAGGTTGGTCTG 20
    ||||| |||||
Db 1764 AGGAAAGGTTGGTCTG 1748

RESULT 11
US-09-657-481A-3/c
; Sequence 3, Application US/09657481A
; Patent No. 6258601
; GENERAL INFORMATION:
; APPLICANT: Brett P. Monia
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF UBIQUITIN PROTEIN LIGASE-WWP1 AND WW
; TITLE OF INVENTION: EXPRESSION
```

```
; FILE REFERENCE: RTS-0087
; CURRENT APPLICATION NUMBER: US/09/657,481A
; CURRENT FILING DATE: 2000-09-07
; NUMBER OF SEQ ID NOS: 93
; SEQ ID NO 3
; LENGTH: 2052
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1).....(2052)
US-09-657-481A-3

Query Match      77.0%; Score 15.4; DB 4; Length 2052;
Best Local Similarity 94.1%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 AGGAAAGGTTGGTCTG 20
    ||||| |||||
Db 1764 AGGAAAGGTTGGTCTG 1748

RESULT 12
US-08-545-809A-32/c
; Sequence 32, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 341 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: human lymphoblast
; CELL LINE: CGM1
US-08-545-809A-32

Query Match      76.0%; Score 15.2; DB 3; Length 341;
Best Local Similarity 85.0%; Pred. No. 32;
```

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20  
||||| ||| |||||

Db 233 TGCAGGAAGTGTGTCTG 214

## RESULT 13

US-08-840-146-16  
; Sequence 16, Application US/08840146  
; Patent No. 6037173

## ; GENERAL INFORMATION:

; APPLICANT: Glucksmann, M. Alexandra

; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS AND

; TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR DISEASES INVOLVING TRBP

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY, HOAG & ELIOT LLP

; STREET: One Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2170

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/840,146

; FILING DATE: 11-APR-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Arnold, Beth E.

; REGISTRATION NUMBER: 35,430

; REFERENCE/DOCKET NUMBER: MIA-018.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000

; TELEFAX: 617-832-7000

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 554 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-840-146-16

Query Match 76.0%; Score 15.2; DB 3; Length 554;

Best Local Similarity 85.0%; Pred. No. 36;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20

||||| ||| |||||

Db 490 TGCAGGACTGTTGGTCTG 509

## RESULT 14

US-09-360-220-16

; Sequence 16, Application US/09360220

; Patent No. 6046308

; GENERAL INFORMATION:

; APPLICANT: Glucksmann, M. Alexandra

; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS AND

; TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR DISEASES INVOLVING TRBP

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY, HOAG & ELIOT LLP

; STREET: One Post Office Square

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109-2170

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/360,220  
; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/840,146

; FILING DATE: 11-APR-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Arnold, Beth E.

; REGISTRATION NUMBER: 35,430

; REFERENCE/DOCKET NUMBER: MIA-018.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-832-1000

; TELEFAX: 617-832-7000

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 554 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-09-360-220-16

Query Match 76.0%; Score 15.2; DB 3; Length 554;

Best Local Similarity 85.0%; Pred. No. 36;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20

||||| ||| |||||

Db 490 TGCAGGACTGTTGGTCTG 509

## RESULT 15

US-08-961-527-325

; Sequence 325, Application US/08961527

; Patent No. 6420135

; GENERAL INFORMATION:

; APPLICANT: Charles Kunsch

; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

; NUMBER OF SEQUENCES: 391

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/961,527

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Brookes, A. Anders

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB340P1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 325:

; SEQUENCE CHARACTERISTICS:

;  
; LENGTH: 1237 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-527-325

Query Match 76.0%; Score 15.2; DB 4; Length 1237;  
Best Local Similarity 85.0%; Pred. No. 43;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCAGGAAGGTTTGGTCTG 20  
|| ||| ||||| |||||  
Db 26 TGGCGGATAGGTTTGGTCTG 45

Search completed: July 8, 2003, 09:32:01  
Job time : 29.6598 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 115.512 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-14  
Perfect score: 20  
Sequence: 1 tgcaggaaggttggtctg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	402	9	US-10-025-380-657
2	16.8	84.0	402	10	US-09-922-217-657
3	16.8	84.0	402	10	US-09-833-263-657
c 4	16.8	84.0	509	9	US-09-918-995-22909
c 5	16.8	84.0	16854	9	US-10-079-854-217
c 6	16.8	84.0	16854	10	US-09-764-878-217
c 7	16.8	84.0	98865	10	US-09-770-689A-3
c 8	16	80.0	656	10	US-09-917-800A-1041
c 9	16	80.0	961	12	US-10-023-335-3
c 10	15.8	79.0	200	10	US-09-864-761-24121
c 11	15.8	79.0	323	9	US-09-796-692-2991
c 12	15.8	79.0	323	9	US-10-040-862-2991
c 13	15.8	79.0	525	9	US-10-016-249-29
c 14	15.8	79.0	525	12	US-10-040-916-29
c 15	15.8	79.0	550	9	US-09-918-995-27372
c 16	15.8	79.0	551	10	US-09-864-761-7458
c 17	15.8	79.0	1212	9	US-10-114-893-157
c 18	15.8	79.0	1464	10	US-09-974-300-32
c 19	15.8	79.0	2000	9	US-09-938-842A-3428

c 20	15.8	79.0	32169	9	US-09-764-891-8605	Sequence 8605, Ap
c 21	15.8	79.0	32189	9	US-09-764-891-8604	Sequence 8604, Ap
c 22	15.8	79.0	34668	9	US-09-900-449A-3	Sequence 3, Appl1
c 23	15.8	79.0	335913	9	US-09-754-853A-2	Sequence 2, Appl1
c 24	15.8	79.0	335913	9	US-09-754-853A-3	Sequence 3, Appl1
c 25	15.4	77.0	410	9	US-09-918-995-8439	Sequence 8439, Ap
c 26	15.4	77.0	452	9	US-09-918-995-32502	Sequence 32502, A
c 27	15.4	77.0	466	9	US-09-854-133-217	Sequence 217, App
c 28	15.4	77.0	466	9	US-10-144-649A-217	Sequence 217, App
c 29	15.4	77.0	466	10	US-09-738-973-217	Sequence 217, App
c 30	15.4	77.0	478	9	US-09-918-995-31836	Sequence 31836, A
c 31	15.4	77.0	506	10	US-09-917-800A-1180	Sequence 1180, Ap
c 32	15.4	77.0	507	9	US-10-025-380-154	Sequence 154, App
c 33	15.4	77.0	507	10	US-09-922-217-154	Sequence 154, App
c 34	15.4	77.0	507	10	US-09-833-263-154	Sequence 154, App
c 35	15.4	77.0	510	9	US-09-918-995-30327	Sequence 30327, A
c 36	15.4	77.0	542	9	US-10-198-846-9425	Sequence 9425, Ap
c 37	15.4	77.0	685	9	US-10-144-929-40	Sequence 40, Appl1
c 38	15.4	77.0	887	9	US-10-036-542-13	Sequence 13, Appl1
c 39	15.4	77.0	897	9	US-10-205-823-41	Sequence 41, Appl1
c 40	15.4	77.0	897	9	US-10-177-293-22	Sequence 22, Appl1
c 41	15.4	77.0	948	10	US-09-764-864-361	Sequence 361, App
c 42	15.4	77.0	968	10	US-09-770-445-291	Sequence 291, App
c 43	15.4	77.0	1034	9	US-10-198-846-14038	Sequence 14038, A
c 44	15.4	77.0	1173	9	US-10-036-542-37	Sequence 37, Appl1
c 45	15.4	77.0	1173	10	US-09-925-301-465	Sequence 465, App

ALIGNMENTS

RESULT 1  
US-10-025-380-657  
; Sequence 657, Application US/10025380  
; Publication No. US20020182191A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yudi  
; APPLICANT: Smith, Carole L.  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Skeiky, Yasir A. W.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedwick Thomas S.  
; APPLICANT: Carter, Darrick  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C14  
; CURRENT APPLICATION NUMBER: US/10/025,380  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 1129  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 657  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 150..153, 154  
; OTHER INFORMATION: n = A,T,C or G  
US-10-025-380-657

Query Match 84.0%; Score 16.8; DB 9; Length 402;  
Best Local Similarity 90.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20  
||||| ||||||||| ||  
Db 366 TGCAGGAAGGTTGGTTG 385

## RESULT 2

US-09-922-217-657  
; Sequence 657, Application US/09922217  
; Patent No. US2002007641A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Meagher, Madeleine Joy  
; APPLICANT: Stolk, John A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Smith, Carole Lynn  
; APPLICANT: King, Gordon E.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C13  
; CURRENT APPLICATION NUMBER: US/09/922,217  
; CURRENT FILING DATE: 2001-08-03  
; NUMBER OF SEQ ID NOS: 1124  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 657  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 150, 153, 154  
; OTHER INFORMATION: n = A,T,C or G  
US-09-922-217-657

Query Match 84.0%; Score 16.8; DB 10; Length 402;  
Best Local Similarity 90.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20  
||||| ||||||||| ||  
Db 366 TGCAGGAAGGTTGGTTG 385

## RESULT 3

US-09-833-263-657  
; Sequence 657, Application US/09833263  
; Patent No. US20020110547A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Aijun  
; APPLICANT: Clapper, Jonathan D.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Meagher, Madeleine J.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.471C12  
; CURRENT APPLICATION NUMBER: US/09/833,263  
; CURRENT FILING DATE: 2001-04-10  
; NUMBER OF SEQ ID NOS: 1093  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 657  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Homo sapien  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(402)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-833-263-657

Query Match 84.0%; Score 16.8; DB 10; Length 402;  
Best Local Similarity 90.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20  
||||| ||||||||| ||  
Db 366 TGCAGGAAGGTTGGTTG 385

## RESULT 4

US-09-918-995-22909/c  
; Sequence 22909, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 22909  
; LENGTH: 509  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(509)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-22909

Query Match 84.0%; Score 16.8; DB 9; Length 509;  
Best Local Similarity 90.0%; Pred. No. 53;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20  
||||| ||||||||| ||  
Db 241 TGCAGGACAGGTTGGTTG 222

## RESULT 5

US-10-079-854-217/c  
; Sequence 217, Application US/10079854  
; Publication No. US20030054368A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA121C1  
; CURRENT APPLICATION NUMBER: US/10/079,854  
; CURRENT FILING DATE: 2002-02-22  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 428  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 217  
; LENGTH: 16854  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-079-854-217

Query Match 84.0%; Score 16.8; DB 9; Length 16854;  
Best Local Similarity 90.0%; Pred. No. 63;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCAGGAAGGTTGGTCTG 20  
||||| ||||||||| ||  
Db 5587 TGCAGGACAGGTTGGTTG 5568

## RESULT 6

US-09-764-878-217/c  
; Sequence 217, Application US/09764878  
; Patent No. US20020090615A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA121  
; CURRENT APPLICATION NUMBER: US/09/764,878  
; PRIOR FILING DATE: 2001-01-17  
; Prior application data removed - consult PAM or file wrapper  
; NUMBER OF SEQ ID NOS: 428  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 217  
; LENGTH: 16854  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-878-217

Query Match 84.0%; Score 16.8; DB 10; Length 16854;  
Best Local Similarity 90.0%; Pred. No. 63;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGTTGGTCTG 20  
||||||| ||||||| ||  
Db 5587 TGCAGGACAGTTGGTTTG 5568

RESULT 7  
US-09-770-689A-3/c  
; Sequence 3, Application US/09770689A  
; Patent No. US20020115171A1  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL001079  
; CURRENT APPLICATION NUMBER: US/09/770,689A  
; CURRENT FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 98865  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-09-770-689A-3

Query Match 84.0%; Score 16.8; DB 10; Length 98865;  
Best Local Similarity 90.0%; Pred. No. 69;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGTTGGTCTG 20  
||||||| ||||||| ||  
Db 85606 TGCAGGACAGTTGGTTTG 85587

RESULT 8  
US-09-917-800A-1041  
; Sequence 1041, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31

; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1041  
; LENGTH: 656  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1178741  
US-09-917-800A-1041

Query Match 80.0%; Score 16; DB 10; Length 656;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 AGGAAAGTTGGTCT 19  
||||||| ||||||| ||  
Db 596 AGGAAAGTTGGTCT 611

RESULT 9  
US-10-023-335-3/c  
; Sequence 3, Application US/10023335  
; Patent No. US20020138856A1  
; GENERAL INFORMATION:  
; APPLICANT: REDEI, Eva  
; APPLICANT: AIRD, Frasier  
; APPLICANT: SOLBERG, Leah  
; APPLICANT: WILL, Claire  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS USEFUL FOR TREATMENT OF DEPRESSIVE D  
; FILE REFERENCE: 053662-5002-01  
; CURRENT APPLICATION NUMBER: US/10/023,335  
; CURRENT FILING DATE: 2001-12-17  
; PRIOR APPLICATION NUMBER: US 09/590,837  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: PCT/US99/17513  
; PRIOR FILING DATE: 1999-08-03  
; PRIOR APPLICATION NUMBER: US 60/105,459  
; PRIOR FILING DATE: 1998-10-23  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 961  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-10-023-335-3

Query Match 80.0%; Score 16; DB 12; Length 961;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCAGGAAAGTTGGT 16  
||||||| ||||||| ||  
Db 166 TGCAGGAAAGTTGGT 151

RESULT 10

US-09-864-761-24121/c  
; Sequence 24121, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 24121  
; LENGTH: 200  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO ALL39347.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.3  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.88  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
; OTHER INFORMATION: EST HUMAN HIT: BF690757.1, EVALUATE 3.00e-49  
; OTHER INFORMATION: NT HIT: g14758789, EVALUATE 2.00e-52  
; OTHER INFORMATION: SWISSPROT HIT: O43920, EVALUATE 3.00e-18  
US-09-864-761-24121

Query Match 79.0%; Score 15.8; DB 10; Length 200;  
Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCAGGAAAGCTTTGGTCT 19  
||||| ||||||| ||  
Db 129 TGCAGAAAGCTTTGGCCT 111  
RESULT 11  
US-09-796-692-2991  
; Sequence 2991, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
; FILE REFERENCE: 2077.001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 9597  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2991  
; LENGTH: 323  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-09-796-692-2991

Query Match 79.0%; Score 15.8; DB 9; Length 323;  
Best Local Similarity 89.5%; Pred. No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GCAGGAAAGCTTTGGTCTG 20  
||||| ||||||| ||  
Db 254 GCAGAAAGCTTTGGTCTG 272

RESULT 12  
US-10-040-862-2991  
; Sequence 2991, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862

;; CURRENT FILING DATE: 2001-11-06  
;; PRIOR APPLICATION NUMBER: US 60/186,126  
;; PRIOR FILING DATE: 2000-03-01  
;; PRIOR APPLICATION NUMBER: US 60/190,479  
;; PRIOR FILING DATE: 2000-03-17  
;; PRIOR APPLICATION NUMBER: US 60/200,545  
;; PRIOR FILING DATE: 2000-04-27  
;; PRIOR APPLICATION NUMBER: US 60/200,303  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,779  
;; PRIOR FILING DATE: 2000-04-28  
;; PRIOR APPLICATION NUMBER: US 60/200,999  
;; PRIOR FILING DATE: 2000-05-01  
;; PRIOR APPLICATION NUMBER: US 60/202,084  
;; PRIOR FILING DATE: 2000-05-04  
;; PRIOR APPLICATION NUMBER: US 60/206,201  
;; PRIOR FILING DATE: 2000-05-22  
;; PRIOR APPLICATION NUMBER: US 60/218,950  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: US 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: US 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; PRIOR APPLICATION NUMBER: US 09/796,692  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 10467  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2991  
;; LENGTH: 323  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-040-862-2991

Query Match 79.0% Score 15.8; DB 9; Length 323;  
Best Local Similarity 89.5%; Pred No. 1.6e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GCAGGAAAGTTTGGTCTG 20  
||||| ||||| ||||| |||||  
Db 254 GCAGAAAGTTTGGTCTG 272

RESULT 13  
US-10-016-249-29/c  
; Sequence 29, Application US/10016249  
; Publication No. US20030100053A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: ML, Sha  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6006B AJ172A  
; CURRENT APPLICATION NUMBER: US/10/016,249  
; CURRENT FILING DATE: 2001-10-30  
; PRIOR APPLICATION NUMBER: US/09/175,928  
; PRIOR FILING DATE: 1998-10-20  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (22)

;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (52)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (55)  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (59)  
US-10-016-249-29

Query Match 79.0% Score 15.8; DB 9; Length 525;  
Best Local Similarity 89.5%; Pred No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GCAGGAAAGTTTGGTCTG 20  
||||| ||||| ||||| |||||  
Db 388 GCAGAAAGCTTGGCTG 370

RESULT 14  
US-10-040-916-29/c  
; Sequence 29, Application US/10040916  
; Patent No. US20020146769A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John  
; APPLICANT: Lavallie, Edward  
; APPLICANT: Racie, Lisa  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Spaulding, Vikki  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; ENCODING THEM  
; NUMBER OF SEQUENCES: 71  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/040,916  
FILING DATE: 07-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/887,029  
FILING DATE: 07-FEB-1997  
APPLICATION NUMBER: 08/686,878  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 525 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 29:  
US-10-040-916-29

Query Match 79.0%; Score 15.8; DB 12; Length 525;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTGGTCTG 20  
| | | | | | | | | | | | | | | | | |  
DB 388 GCAGGAAGGTTGGGCTG 370

## RESULT 15

US-09-918-995-27372/c  
; Sequence 27372, Application US/0918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 27372  
; LENGTH: 550  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(550)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-27372

Query Match 79.0%; Score 15.8; DB 9; Length 550;  
Best Local Similarity 89.5%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGAGGAAGGTTGGTCT 19  
| | | | | | | | | | | | | | | | | |  
DB 215 TTCAGGAAGGTTGGTCT 197

Search completed: July 9, 2003, 02:22:19  
Job time : 115.562 secs

pred. No. is the number of results predicted by chance to have a

## SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	PB		
1	20	100.0	20	6	AX014714	AX014714 Sequence
2	20	100.0	2765	6	AX014701	AX014701 Sequence
C 3	20	100.0	174953	2	AC094732	AC094732 Rattus no
C 4	18.4	92.0	336	11	G66449	G66449 s12509 YAC
C 5	18.4	92.0	1529	9	AR169782	AR169782 Sequence
C 6	18.4	92.0	1870	9	BC025363	BC025363 Homo sapi
C 7	18.4	92.0	2061	9	AB031038	AB031038 Homo sapi
C 8	18.4	92.0	2494	6	AR169779	AR169779 Sequence
C 9	18.4	92.0	2583	10	AF302075	AF302075 Mus muscu
10	18.4	92.0	2601	10	AF157106	AF157106 Mus muscu
11	18.4	92.0	2652	10	AF302076	AF302076 Mus muscu
12	18.4	92.0	2694	10	AF302077	AF302077 Mus muscu
13	18.4	92.0	2892	10	AF157105	AF157105 Mus muscu
14	18.4	92.0	2925	6	AX033272	AX033272 Sequence
15	18.4	92.0	2925	10	AF176569	AF176569 Mus muscu
16	18.4	92.0	163820	2	AC123878	AC123878 Mus muscu
17	18.4	92.0	200379	9	AC098614	AC098614 Homo sapi
18	18.4	92.0	208249	2	AL607032	AL607032 Mus muscu
C 19	18.4	92.0	216172	9	AC098973	AC098973 Homo sapi
C 20	17.4	87.0	3323	6	A90770	A90770 Sequence 5
C 21	17.4	87.0	3323	6	A90771	A90771 Sequence 6
22	17.4	87.0	3323	8	SCYOR005C	274913 S.cerevisia
23	17.4	87.0	30469	8	SCU34391	043491 Saccharomyc
C 24	17.4	87.0	36016	9	HSUDCA10	275742 Human DNA s
C 25	17.4	87.0	38275	2	AC069372	AC069372 Homo sapi
C 26	17.4	87.0	71295	2	AC124135	AC124135 Mus muscu
C 27	17.4	87.0	71295	2	AC124135	AC124135 Mus muscu
28	17.4	87.0	148466	2	AC023167	AC023167 Mus muscu
29	17.4	87.0	157496	2	AC111520	AC111520 Rattus no
30	17.4	87.0	165337	2	AC096212	AC096212 Rattus no
C 31	17.4	87.0	167061	9	AC092380	AC092380 Homo sapi
C 32	17.4	87.0	173294	2	AC105317	AC105317 Homo sapi
C 33	17.4	87.0	179380	2	AC120325	AC120325 Rattus no
C 34	17.4	87.0	179999	2	AC023427	AC023427 Mus muscu
35	17.4	87.0	182490	2	AC068572	AC068572 Homo sapi
36	17.4	87.0	184451	9	AC016768	AC016768 Homo sapi
C 37	17.4	87.0	185293	2	AC126585	AC126585 Rattus no
C 38	17.4	87.0	185829	2	AC094911	AC094911 Rattus no
39	17.4	87.0	194590	2	AL645962	AL645962 Mus muscu
40	17.4	87.0	201467	2	AC106371	AC106371 Rattus no
41	17.4	87.0	207512	2	AC116236	AC116236 Rattus no
42	17.4	87.0	209639	10	AC079438	AC079438 Mus muscu
43	17.4	87.0	216106	10	AL627215	AL627215 Mouse DNA
44	17.4	87.0	2259159	2	AC124756	AC124756 Mus muscu
45	17	85.0	326	9	HS2NGP2P3	X70171 H.sapiens D

## ALLEGMENTS

[illegible]

JOURNAL Patent: WO 9953077-A 14 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUIMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES  
source  
1. 20  
Location/Qualifiers  
1. 20  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="oligonucleotide"  
4 a 2 c 8 g 6 t  
BASE COUNT  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCAGGAAGTTTGGTCTG 20  
- - - - -  
Db 1 TGCAGGAAGTTTGGTCTG 20  
RESULT 2  
AX014701  
LOCUS AX014701 2765 bp. DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 1 from Patent W09953077.  
ACCESSION AX014701  
VERSION AX014701.1 GI:10040975  
KEYWORDS black rat.  
SOURCE Rattus rattus  
ORGANISM Rattus rattus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 2765)  
AUTHORS Schwartz,J.C., Gros,C., Ouimet,T., Rose,C., Bonhomme,M.C. and  
Facchinetti,P.  
TITLE Novel hep ii membrane metalloprotease and its use for screening,  
inhibitors useful in therapy  
JOURNAL Patent: WO 9953077-A 1 21-OCT-1999;  
INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
CLAUDE (FR); OUIMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)  
FEATURES  
source  
1. 2765  
/organism="Rattus rattus"  
/db\_xref="taxon:10117"  
107. 2431  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07576.1"  
/db\_xref="GI:10040976"  
/translation="MGKSSVGMERADNCGRRRIGFVPCGLLVLLTLLMGIVTL  
GVPTSGKQLPLNLLSHVSRHTVVRVLRDSSQKSDICTTPSCVIAARILQNM  
QSKPCDFQYACGGVHRHVLPETNSYSVFDILRDEVLKGLVDESSVQHRPA  
VEKATLYRSCNMQVIEKRDSEPLNLVDMIGVPMQKMDKWTMGPKWELERQAV  
LNSQFNRRVLDLFIWDDQNSRRHVIYIDPTLGMPSREYFKEDSHRYREAYLQEM  
TSVATMLRDLNLPGETDLYQSEMAQVHLHETLANATVPQEKHVDYLYHMGLEE  
LQRFGLKGNWTLFTQNLVSSVQVELLPNEEVVYIGPILENLEIIDVFPAGTLLQ  
YLWRLVLDIGLSQRFKEARVDYRKALYGTMEEVVRRECVSYVNSMESAVGSY  
IKRAFSDKDSIVSELEKIRSVFDNLDELNMDESKKKAQEKALNIREQIGPDY  
ILEDNRNKLIVSESTFSDLYFENGLQNKNAQSLKRLKREKVDONLIGAAVY  
NAPYSRNRLNIVFPAGTLPQPFPSKQPOALNFGGIGWIGHEITHFGDDNGRNFKN  
GNLMDWSNFSARFRQOSOCILYQYSNFSWELADNONGVSTGLGNIDNGVROA  
YKAYLWLAGBGDQRLPGLNLYIAQLFFINAYQVWCGSVRPFAIOSIKYDVHSPLK  
IRVIGLSQNLPGFSRFAHCPGSGPMHNNCRILW"  
BASE COUNT 684 a 735 c 787 g 559 t  
ORIGIN  
Query Match 100.0%; Score 20; DB 6; Length 2765;  
Best Local Similarity 100.0%; Pred. No. 5.6;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCAGGAAGTTTGGTCTG 20

Db 1110 TGCAGGAAGTTTGGTCTG 1129

RESULT 3  
AC094732/c  
LOCUS AC094732 174953 bp DNA linear HTG 20-DEC-2001  
DEFINITION Rattus norvegicus clone CH230-516, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
56 unordered pieces.  
ACCESSION AC094732.2 GI:17941511  
VERSION HTG; HTGS\_PHASE1.  
KEYWORDS Rattus norvegicus.  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 174953)  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,  
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,  
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,  
Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Kratovick,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,  
Louisseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,B., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,  
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,  
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,  
Sisson,I., Sodergren,E., Sonaika,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 174953)  
Direct Submission  
Worley,K.C.  
JOURNAL  
TITLE  
JOURNAL  
COMMENT  
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Dec 20, 2001 this sequence version replaced gi:15624568.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GBGF

Center clone name: CH230-5f6  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329First call to findPhrapList  
 Consensus quality: 152255 bases at least Q40  
 Consensus quality: 158448 bases at least Q30  
 Consensus quality: 164461 bases at least Q20  
 Estimated insert size: 155965; sum-of-contigs estimation  
 Quality coverage: 0x in Q20 bases; agarose-gel estimation  
 Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 56 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved

1 13782: contig of 13782 bp in length  
 \* 13783 13882: gap of unknown length  
 \* 13883 23287: contig of 9405 bp in length  
 \* 23288 23387: gap of unknown length  
 \* 23388 28081: contig of 4694 bp in length  
 \* 28082 28181: gap of unknown length  
 \* 28182 33807: contig of 5626 bp in length  
 \* 33808 33907: gap of unknown length  
 \* 33908 39271: contig of 5364 bp in length  
 \* 39272 39371: gap of unknown length  
 \* 39372 44270: contig of 4899 bp in length  
 \* 44271 44370: gap of unknown length  
 \* 44371 47723: contig of 3353 bp in length  
 \* 47724 47823: gap of unknown length  
 \* 47824 53427: contig of 5604 bp in length  
 \* 53428 53527: gap of unknown length  
 \* 53528 57303: contig of 3776 bp in length  
 \* 57304 57403: gap of unknown length  
 \* 57404 62018: contig of 4615 bp in length  
 \* 62019 62118: gap of unknown length  
 \* 62119 66676: contig of 4558 bp in length  
 \* 66677 66776: gap of unknown length  
 \* 66777 70201: contig of 3425 bp in length  
 \* 70202 70301: gap of unknown length  
 \* 70302 74082: contig of 3781 bp in length  
 \* 74083 74182: gap of unknown length  
 \* 74183 78306: contig of 4124 bp in length  
 \* 78307 78406: gap of unknown length  
 \* 78407 81222: contig of 2816 bp in length  
 \* 81223 81322: gap of unknown length  
 \* 81323 84552: contig of 3230 bp in length  
 \* 84553 84652: gap of unknown length  
 \* 84653 88839: contig of 4187 bp in length  
 \* 88840 88939: gap of unknown length  
 \* 88940 92256: contig of 3317 bp in length  
 \* 92257 92356: gap of unknown length  
 \* 92357 94995: contig of 2639 bp in length  
 \* 94996 95095: gap of unknown length  
 \* 95096 98465: contig of 3370 bp in length  
 \* 98466 98565: gap of unknown length  
 \* 98566 101632: contig of 3067 bp in length  
 \* 101633 101732: gap of unknown length  
 \* 101733 104939: contig of 3207 bp in length  
 \* 104940 105039: gap of unknown length  
 \* 105040 109248: contig of 4209 bp in length  
 \* 109249 109348: gap of unknown length  
 \* 109349 112435: contig of 3087 bp in length  
 \* 112436 112535: gap of unknown length  
 \* 112536 114676: contig of 2141 bp in length  
 \* 114677 114776: gap of unknown length  
 \* 114777 117567: contig of 2791 bp in length  
 \* 117568 117667: gap of unknown length

\* 117668 119209: contig of 1542 bp in length  
 \* 119210 119309: gap of unknown length  
 \* 119310 121797: contig of 2488 bp in length  
 \* 121798 121897: gap of unknown length  
 \* 121898 124293: contig of 2396 bp in length  
 \* 124294 124393: gap of unknown length  
 \* 124394 126438: contig of 2035 bp in length  
 \* 126429 126528: gap of unknown length  
 \* 126529 129525: contig of 2997 bp in length  
 \* 129526 129625: gap of unknown length  
 \* 129626 132334: contig of 2709 bp in length  
 \* 132335 132434: gap of unknown length  
 \* 132435 135274: contig of 2840 bp in length  
 \* 135275 135374: gap of unknown length  
 \* 135375 138074: contig of 2700 bp in length  
 \* 138075 138174: gap of unknown length  
 \* 138175 139985: contig of 1811 bp in length  
 \* 139986 140085: gap of unknown length  
 \* 140086 142273: contig of 2188 bp in length  
 \* 142274 142373: gap of unknown length  
 \* 142374 143598: contig of 1225 bp in length  
 \* 143599 143698: gap of unknown length  
 \* 143699 145434: contig of 1736 bp in length  
 \* 145435 145534: gap of unknown length  
 \* 145535 146985: contig of 1451 bp in length  
 \* 146986 147085: gap of unknown length  
 \* 147086 148099: contig of 1014 bp in length  
 \* 148100 148199: gap of unknown length  
 \* 148200 150915: contig of 2716 bp in length  
 \* 150916 151015: gap of unknown length  
 \* 151016 152501: contig of 1486 bp in length  
 \* 152502 152601: gap of unknown length  
 \* 152602 154010: contig of 1409 bp in length  
 \* 154011 154111: gap of unknown length  
 \* 154112 155758: contig of 1648 bp in length  
 \* 155759 155858: gap of unknown length  
 \* 155859 157622: contig of 1764 bp in length  
 \* 157623 157722: gap of unknown length  
 \* 157723 159428: contig of 1706 bp in length  
 \* 159429 159528: gap of unknown length  
 \* 159529 161209: contig of 1681 bp in length  
 \* 161210 161309: gap of unknown length  
 \* 161310 163413: contig of 2104 bp in length  
 \* 163414 163513: gap of unknown length  
 \* 163514 164702: contig of 1189 bp in length  
 \* 164703 164802: gap of unknown length  
 \* 164803 165998: contig of 1196 bp in length  
 \* 165999 166098: gap of unknown length  
 \* 166099 167412: contig of 1314 bp in length  
 \* 167413 167512: gap of unknown length  
 \* 167513 169231: contig of 1719 bp in length  
 \* 169232 169331: gap of unknown length  
 \* 169332 170534: contig of 1203 bp in length  
 \* 170535 170634: gap of unknown length  
 \* 170635 172047: contig of 1413 bp in length  
 \* 172048 172147: gap of unknown length  
 \* 172148 173509: contig of 1362 bp in length  
 \* 173510 173609: gap of unknown length  
 \* 173610 174953: contig of 1344 bp in length.  
 \* Location/Qualifiers

## FEATURES

Query Match 100.0%; Score 20; DB 2; Length 174953;  
 Best Local Similarity 100.0%; Pred. No. 7.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCAGGAAAGGTTTGGTCTG 20

DB 58710 TCCAGGAAAGGTTTGGTCTG 58691

## RESULT 4

G66449/c

LOCUS

G66449

396 bp

DNA

linear

STS 02-NOV-2001

DEFINITION sv2509 YAC Subtraction Homo sapiens STS genomic, sequence tagged site.

ACCESSION G66449

VERSION G66449.1 GI:15078500

KEYWORDS STS.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Tilford, C.A., Rosen, B., Kuroda-Kawaguchi, T., Skaletsky, H., Rozen, S., Brown, L.G., Kosenberg, M., McPherson, J.D., Wyllie, K., Sekhon, M., Kucaba, T.A., Waterston, R.H., and Page, D.C.

TITLE A physical map of the human Y chromosome

JOURNAL Nature 409 (6822), 943-945 (2001)

MEDLINE 21131744

PUBMED 11237016

COMMENT Synonyms: SP37-36

Contact: Charles A. Tilford

Page Lab Whitehead Institute for Biomedical Research

Room 425, 9 Cambridge Center, Cambridge, MA 02142, USA

Email: tilford@wi.mit.edu

Primer A: TACGGCTTCAATCTTCTT

Primer B: CACGGTCTCTCGCCATAT

STS size: 81

PCR Profile:

94C	4:00 min
/ 94C	:30 sec
35x / 58C	2:00 min
/ 72C	2:00 min
72C	7:00 min

Protocol:

Template:	25-100 ng
Primer:	each 1 uM
dNTPs:	each 100 uM
Taq Polymerase:	0.05 units/uL
Total Vol:	15 uL
Mineral Oil:	15 uL

Buffer: For 20ml of 10x Stock solution:

Stock Reagent	Volume	Mixed
1M KCl	10 ml	
1M Tris-HCl, pH 9.0	2 ml	
Triton X-100	200 uL	
2M MgCl2	150 uL	
H2O	7.65 ml	

Final Conc. of 10x Stock

500 mM
100 mM
1.0 %
15 mM

FEATURES

source

1..396

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/sex="male"

/cell\_line="OXEN XYXY Male"

/clone\_lib="YAC Subtraction"

/note="Vector: pYAC4; V-type: YAC; DNA subtraction between Y chromosome YACs used to generate a pool of fragments enriched for low-copy number, Y-specific sequences."

30..110

primer\_bind 30..49

primer\_bind complement(91..110)

BASE COUNT 98 a 124 c 78 g 89 t 7 others

ORIGIN

Query Match 92.0%; Score 18.4; DB 11; Length 396;

Best Local Similarity 95.0%; Pred. No. 38;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAGGAAAGGTTTGGTCTG 20

|||||

Db 337 TCGAGGAAAGGTTTGGTCTG 318

RESULT 5

AR169782/c

LOCUS 1529 bp DNA linear PAT 17-DEC-2001

DEFINITION Sequence 5 from patent US 6291193.

ACCESSION AR169782

VERSION AR169782.1 GI:17907690

KEYWORDS

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1529)

AUTHORS Khodadoust, M.

TITLE MTbx protein and nucleic acid molecules and uses therefor

JOURNAL Patent: US 6291193-A 5 18-SEP-2001.

FEATURES

source

1..1529

/organism="unknown"

BASE COUNT 432 a 342 c 335 g 416 t 4 others

ORIGIN

Query Match 92.0%; Score 18.4; DB 6; Length 1529;

Best Local Similarity 95.0%; Pred. No. 42;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGAGGAAAGGTTTGGTCTG 20

|||||

Db 404 TCGAGGAAAGGTTTGGTCTG 385

RESULT 6

BC025363/c

LOCUS 1870 bp mRNA linear PRI 08-MAR-2002

DEFINITION Homo sapiens, Similar to eomesodermin homolog (Xenopus laevis), clone IMAGE:5013504, mRNA, partial cds.

ACCESSION BC025363

VERSION BC025363.1 GI:19263736

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1870)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (05-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgaps-f@mail.nih.gov](mailto:cgaps-f@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada

info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 42 Row: 1 Column: 3.

FEATURES

source

1..1870

/organism="Homo sapiens"

/db\_xref="taxon:9606"

```

/clone="IMAGE:5013504"
/tissue_type="Brain, astrocytoma, grade IV"
/clone_lib="NIH_MGC_98"
/lab_host="DH108"
/notes="vector: pOTB7"
<1..1175
/codon_start=3
/product="Similar to eomesodermin homolog (Xenopus laevis)"
/protein_id="AAH25363.1"
/db_xref="GI:19263737"
/translation="FFLSNFNLNPTAHYVVEVLADPNHWRFGGKWTGCGA
RNMQCNKMTVHPEPNTSGSHMRQEIISFGKLLTNKNGANNNTOMIVLQSLHKYOP
DLHIVETDVEDLNEPSKTQTFSETQFIATVAYQNTDITQKIDHNPFAKGRD
NYDSHQIIPGGYGVQSPFPFVNTLPQARYNGERTVPOINGLLSPQSEYANP
PQWLVTPOQNTKLDISSYSEYESTSLLPIGKSLPQSHALGYTPDPFFPAM
AGWGGVGSYQKMAAGLPTSTPTVFSDELQSKKEKVEEISWTETPPSYKSLDS
NDSGVYSACKRRRLSPSNSNSNSIKEDINAEYSKDTSKGMGGYAFVTFP"
BASE COUNT      545 a  441 c  402 g  482 t
ORIGIN
Query Match      92.0%; Score 18.4; DB 9; Length 1870;
Best Local Similarity 95.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

-QY 1 TGCAGGAAAGGTTGGTCTG 20
      |||||
Db 827 TGCAGGAAAGGTTGGTCTG 808

RESULT 7
AB031038/c
LOCUS AB031038 2061 bp mRNA linear PRI 19-AUG-1999
DEFINITION Homo sapiens mRNA for hTbr2, complete cds.
ACCESSION AB031038
VERSION AB031038.1 GI:5738951
KEYWORDS hTbr2.
SOURCE Homo sapiens tissue_lib:fetal brain cdna to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (sites)
AUTHORS Kimura,N., Nakashima,K., Ueno,M., Kiyama,H. and Taga,T.
TITLE A novel mammalian T-box-containing gene, Tbr2, expressed in mouse
developing brain
JOURNAL Brain Res. Dev. Brain Res. 115 (2), 183-193 (1999)
MEDLINE 99337662
REFERENCE 2 (bases 1 to 2061)
AUTHORS Kimura,N., Taga,T., Nakashima,K. and Ueno,M.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Naoki Kimura, Tokyo Medical and Dental
University, Department of Molecular Cell Biology, Medical Research
Institute, Kanda-Surugadai 2-3-10, Chiyoda-ku, Tokyo 101-0062,
Japan (E-mail:Kimukbenr1.tmd.ac.jp, Tel:+81-3-5280-8063,
Fax:+81-3-5280-8065)
FEATURES
Location/Qualifiers
1..2061
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="fetal brain"
1..2061
/gene="hTbr2"
1..2061
/gene="hTbr2"
/notes="T-box-containing gene"
/codon_start=1
/product="hTbr2"
/protein_id="BAA83417.1"
/db_xref="GI:5738952"
/translation="MOLGEQLLVSNLPGAHPYLPESARGSGGAGHLPSPAQSPQ
KIDLGKSGKSGSLSCSAVSGEPAAPAAALMSDITDAGDAFASAAVAKPGPPD
GRKSGKSGEELPSRAAAAAAATAARYSMDSLSVYLYLQSPGQSGSLAACP
SLFTQAAAGAPGHPVYPAPNGARYIPGSMPLPPGPPFAAACPPGAPGAGAGCA

```

```

GGINGGGGPGCTYOSGAPLYGPYGAAGSCGGLGGLGVPPSGFRAHVYLCNRPL
WLKFRHQTMLITKQGRPEPFLSNFNLNPTAHYVVEVLADPNHWRFGGKWK
VTCGKADNMCGNKMVHPSPNGSHMRQEIISFGKLLTNKNGANNNTOMIVLQSP
LHKYOPRHLHIVEETDVEDLNEPSKTQTFSETQFIATVAYQNTDITQKIDHNP
AKGRDNTDSSHQIIPGGYGVQSPFPFVNTLPQARYNGERTVPOINGLLSPQSP
EEVANPPQRLWLTVPQPGTNKLDISSYSEYESTSLLPIGKSLPQSHALGYTPD
PTFPAMAGWGGRGSGYQKMAAGLPTSTPTVFSDELQSKKEKVEEISGSSWIEPPS
IKSLDSNDSNGVYTSACKRRRLSPSNSNSNSIKEDINAEYSKDTSKGMGGYAF
YTSP"
BASE COUNT      465 a  640 c  557 g  399 t
ORIGIN
Query Match      92.0%; Score 18.4; DB 9; Length 2061;
Best Local Similarity 95.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

-QY 1 TGCAGGAAAGGTTGGTCTG 20
      |||||
Db 1713 TGCAGGAAAGGTTGGTCTG 1694

RESULT 8
AR169779/c
LOCUS AR169779 2494 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 1 from patent US 6291193.
ACCESSION AR169779
VERSION AR169779.1 GI:17907687
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2494)
AUTHORS Khodadoust,M.
TITLE Mtbx protein and nucleic acid molecules and uses therefor
JOURNAL Patent: US 6291193-A 1 18-SEP-2001;
FEATURES Location/Qualifiers
source 1..2494
/organism="unknown"
BASE COUNT      650 a  639 c  601 g  594 t  10 others
ORIGIN
Query Match      92.0%; Score 18.4; DB 6; Length 2494;
Best Local Similarity 95.0%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

-QY 1 TGCAGGAAAGGTTGGTCTG 20
      |||||
Db 1369 TGCAGGAAAGGTTGGTCTG 1350

RESULT 9
AF302075
LOCUS AF302075 2583 bp mRNA linear ROD 11-JUN-2001
DEFINITION Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.
ACCESSION AF302075
VERSION AF302075.1 GI:10505359
KEYWORDS Mus musculus.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2583)
AUTHORS Shirotani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,
Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,
Iwatsubo,T. and Saido,T.C.
TITLE Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most
rapidly and efficiently among thiorphan- and
phosphoramidon-sensitive endopeptidases
J. Biol. Chem. 276 (24), 21895-21901 (2001).
JOURNAL 21293028
MEDLINE 11278416
PUBMED
REFERENCE 2 (bases 1 to 2583)
AUTHORS Shirotani,K. and Saido,T.C.

```

[illegible]



```

TITLE      Direct Submission
JOURNAL    Submitted (09-AUG-1999) Biochemistry, University of Montreal, 2900
           Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES   Location/Qualifiers
           1..2925
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /tissue_type="testis"
           1..2925
             /gene="Nl1"
             /gene="Nl1"
             /note="Nl1: metallo-endopeptidase"
             /codon_start=1
             /product="neprilysin-like metallopeptidase 1"
             /protein_id="AAF69247.1"
           gene
           CDS

```

[illegible]

Search completed: July 8, 2003, 03:34:53  
Job time : 224.098 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1013.95 Seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-14

Perfect score: 20

Sequence: 1 tgcaggaaagtttggctcg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_estc:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_plp:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	18.4	92.0	311	9	AA323074
c 2	18.4	92.0	364	10	BE679769
c 3	18.4	92.0	608	10	AW965242
c 4	17.4	87.0	355	12	AW965242
c 5	17.4	87.0	385	13	BE413980
c 6	17.4	87.0	390	10	BE104129
					BE104129
					BE192046
					BE192046

7	17.4	87.0	419	12	BF072189
8	17.4	87.0	421	10	BE491715
c 9	17.4	87.0	431	10	AW633486
10	17.4	87.0	434	17	AW633486
c 11	17.4	87.0	453	10	AW633397
c 12	17.4	87.0	455	12	BG016217
c 13	17.4	87.0	463	12	BG016230
c 14	17.4	87.0	475	12	BG016230
c 15	17.4	87.0	488	12	BG016230
c 16	17.4	87.0	503	13	BM507586
c 17	17.4	87.0	552	9	AJ449141
c 18	17.4	87.0	596	14	BQ398336
c 19	17.4	87.0	607	17	AQ391007
c 20	17.4	87.0	627	17	AQ920045
c 21	16.8	84.0	311	10	BB471270
c 22	16.8	84.0	386	12	BG145822
c 23	16.8	84.0	405	13	BJ496701
c 24	16.8	84.0	406	12	BF380110
c 25	16.8	84.0	413	9	AA881740
c 26	16.8	84.0	414	14	BQ901909
c 27	16.8	84.0	417	13	BJ496273
c 28	16.8	84.0	423	12	BG687908
c 29	16.8	84.0	433	9	AI072298
c 30	16.8	84.0	456	9	AI718362
c 31	16.8	84.0	473	9	AI070598
c 32	16.8	84.0	494	14	WA5913
c 33	16.8	84.0	497	17	AQ283623
c 34	16.8	84.0	497	17	AQ597562
c 35	16.8	84.0	502	13	BJ320240
c 36	16.8	84.0	512	17	AZ850482
c 37	16.8	84.0	516	13	BI081952
c 38	16.8	84.0	536	13	BJ500523
c 39	16.8	84.0	549	9	AA874322
c 40	16.8	84.0	588	12	BG084027
c 41	16.8	84.0	600	12	BG804027
c 42	16.8	84.0	604	10	BE304168
c 43	16.8	84.0	615	13	BJ499618
c 44	16.8	84.0	624	13	BJ495288
c 45	16.8	84.0	625	10	BE306772

## ALIGNMENTS

RESULT 1  
AA323074/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS

311 bp mRNA linear EST 20-APR-1997  
EST25804 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.  
AA323074.1 GI:1975399  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 311)  
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult  
C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White  
O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,  
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald  
L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A.,  
Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M.,  
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,  
Moreno-Palancas,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R.,  
Small,R.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,  
Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Fertie,A., Fischer,C., Hastings,G.A., He,W.W.,  
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,  
Kunsch,C., Hunglun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,  
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon  
M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and  
Venter,J.C.

**TITLE** Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

**JOURNAL** Nature 377 (6547 Suppl.), 3-174 (1995)

**MEDLINE** 96026280

**COMMENT** Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

# FEATURES

Location/Qualifiers

1..311  
/organism="Homo sapiens"  
/db\_xref="ATCC (inhost):123602"  
/db\_xref="taxon:9606"  
/clone\_lib="cerebellum II"  
/tissue\_type="cerebellum"  
/dev\_stage="adult"  
/note="Organ: brain; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

**BASE COUNT** 87 a 85 c 65 g 71 t 3 others

**Query Match** 92.0%; Score 18.4; DB 9; Length 311;  
**Best Local Similarity** 95.0%; Pred. No. 1.2e+02;  
**Matches** 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 1 TGCAGGAAGGTTGGTCTG 20  
|||||  
**Db** 143 TGCAGGAAGGTTGGTCTG 124

**RESULT 2**  
**LOCUS** BE679769/c 364 bp mRNA linear EST 31-MAY-2002  
**DEFINITION** df56h10.y1 Xenopus laevis unfertilized egg cDNA library xenopus laevis cDNA clone XENOPUS\_SOURCE\_ID:xlneq002120 5', mRNA sequence.

**ACCESSION** BE679769  
**VERSION** BE679769.1 GI:10062952  
**KEYWORDS** EST:  
**SOURCE** African clawed frog.  
**ORGANISM** Xenopus laevis

**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.  
1 (bases 1 to 364)  
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hallier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R. and Wilson, R.

**TITLE** WashU Xenopus EST project, 1999  
**JOURNAL** Unpublished (1999)  
**COMMENT** Other\_ESTs: df56h10.xl  
Contact: Sandy Clifton, Ph.D.  
WashU Xenopus EST project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Library constructed by Bruce Blumberg  
Library normalized by Jihwan Song  
DNA Sequencing by: Washington University Genome Sequencing Center  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 359.  
Location/Qualifiers

# FEATURES

source

1..364  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="XENOPUS\_SOURCE\_ID:xlneq002120"  
/clone\_lib="Xenopus laevis unfertilized egg cDNA library"  
/tissue\_type="unfertilized egg"  
/lab\_host="Top-10 F"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; This library has been used successfully to clone a number of full-length cDNAs ranging in size from 1.4 to 4.5 kb. There are less than 0.5% clones with multiple inserts. Since each cDNA has an EcoRI site at its 5' end and an XhoI site at the 3' end, these clones can be easily identified. One should be suspicious of any clone which gives 3 or more bands in an EcoRI-XhoI double digest AND has an internal XhoI site. We usually do not further characterize any such clones unless the cDNA is known to give multiple bands in an EcoRI-XhoI digest. Microplate status: 500,000 unamplified cDNAs were mass excised (pBluescript SK-) in Xil-Blue using ExAssist phage. The resulting single-stranded phagemids were used to infect Top10F. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4\*7H2O, 6.8 mM (NH4)2SO4, 4% w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Note: This is a Xenopus Gene Collection (XGC) library"

**BASE COUNT** 132 a 91 c 81 g 60 t

**Query Match** 92.0%; Score 18.4; DB 10; Length 364;  
**Best Local Similarity** 95.0%; Pred. No. 1.3e+02;  
**Matches** 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

**QY** 1 TGCAGGAAGGTTGGTCTG 20  
|||||  
**Db** 241 TGCATGAAGGTTGGTCTG 222

**RESULT 3**  
**LOCUS** AW965242/c 608 bp mRNA linear EST 01-JUN-2000  
**DEFINITION** EST377315 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.  
**ACCESSION** AW965242  
**VERSION** AW965242.1 GI:8155078  
**KEYWORDS** EST:  
**SOURCE** human.  
**ORGANISM** Homo sapiens

**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 608)  
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.

**TITLE** Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: [johnq@tigr.org](mailto:johnq@tigr.org)  
Plate: 212  
Seq primer: Reverse.

# FEATURES

Location/Qualifiers

1..608  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGI"  
/note="Vector: pBluescriptSKm"

```

BASE COUNT      165 a   179 c   120 g   141 t   3 others
ORIGIN

Query Match      92.0%; Score 18.4; DB 10; Length 608;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAGGTTTGCTG 20
    |||||
Db 539 TGCAGGAAGGTTTGCTG 520
    |||||

RESULT 4
BG413980
LOCUS
DEFINITION de76c11.x1 Kirschner embryo St10 14 Xenopus laevis cDNA clone
IMAGE:3517461 3', mRNA sequence.
ACCESSION BG413980
VERSION
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 355)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterson,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by M. Kirschner (Harvard Medical School). DNA
Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 322.
FEATURES
Location/Qualifiers
1..355
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:3517461"
/clone_lib="Kirschner embryo St10 14"
/tissue_type="pooled embryos (stage 10-14)"
/lab_host="DH10B (phage-resistant)"
/note="vector: pCS2+; Site_1: NotI; Site_2: SalI;
Size-selected for average insert size 1.2 kb. Library was
constructed and donated by M. Kirschner (Harvard Medical
School)."
BASE COUNT      38 a   68 c   150 g   99 t
ORIGIN

Query Match      87.0%; Score 17.4; DB 12; Length 355;
Best Local Similarity 94.7%; Pred. No. 3.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTTGCTG 20
    |||||
Db 60 GCTGGAAGGTTTGCTG 78
    |||||

RESULT 5
BG41429
LOCUS
DEFINITION PMO-NT0911-280201-017-f07 NT0911 Homo sapiens cDNA, mRNA sequence.
IMAGE:3375919 3', mRNA sequence.
ACCESSION BE192046
VERSION
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 390)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Pape,D.

```

```

ACCESSION BI041429
VERSION BI041429.1
KEYWORDS GI:14448055
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 385)
Dias Neto,E., Garcia Corrae,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,A.G., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
JOURNAL
MEDLINE
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PMO&t2=PMO-NT0911-
280201-017-f07&t3=2001-02-28&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 385.
High quality sequence stop: 385.
FEATURES
Location/Qualifiers
1..385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0911"
/dev_stage="Adult"
/note="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      121 a   73 c   65 g   126 t
ORIGIN

Query Match      87.0%; Score 17.4; DB 13; Length 385;
Best Local Similarity 94.7%; Pred. No. 4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAGGTTTGCTCT 19
    |||||
Db 141 TGAAGGAAGGTTTGCTCT 159
    |||||

RESULT 6
BE192046
LOCUS
DEFINITION BE192046.1 Wellcome CRC pSK animal cap-Xenopus laevis cDNA clone
IMAGE:3375919 3', mRNA sequence.
ACCESSION BE192046
VERSION
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 390)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Pape,D.

```

**TITLE**  
**JOURNAL**  
**COMMENT**

Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R., and Willson, R.  
 WashU Xenopus EST project, 1999  
 Unpublished (1999)  
 On Jun 23, 2000 this sequence version replaced gi:8672410.  
 Other ESTs: db86f12.y1  
 Contact: Sandy Clifton, Ph.D.  
 WashU Xenopus EST project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
 Seq primer: -40UP from Gibco.

**FEATURES**  
 source

1..390  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:3379919"  
 /clone\_lib="Wellcome CRC psk animal cap"  
 /tissue\_type="embryo, animal cap"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Vector: pBluescript SK-; Site\_1: NotI; Site\_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library was constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). Note: This is a Xenopus Gene Collection (XGC) library."  
 140 a 59 c 91 g 100 t

**BASE COUNT**  
**ORIGIN**

Query Match 87.0%; Score 17.4; DB 10; Length 390;  
 Best Local Similarity 94.7%; Pred. No. 4e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GCAGGAAAGGTTTGGTCTG 20  
 |||||  
 Db 24 GCTGAAAGGTTTGGTCTG 42

**RESULT 7**

**LOCUS**  
**DEFINITION**  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**

BF072189  
 db52h07.x1 Blackshear/Soares normalized xenopus egg library  
 laevis cDNA clone IMAGE:3302173 3', mRNA sequence.  
 BF072189  
 BF072189.1 GI:10848828  
 EST.  
 African clawed frog.  
 Xenopus laevis

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

1 (bases 1 to 419)  
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R., and Willson, R.  
 WashU Xenopus EST project, 1999  
 Unpublished (1999)  
 Contact: Sandy Clifton, Ph.D.  
 WashU Xenopus EST project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Library constructed by Bento Soares and M. Fatima Bonaldo (University of Iowa). DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
 Seq primer: -40UP from Gibco.

**FEATURES**  
 source

1..419  
 /organism="Xenopus laevis"  
 /db\_xref="taxon:8355"  
 /clone="IMAGE:3302173"  
 /clone\_lib="Blackshear/Soares normalized Xenopus egg library"  
 /sex="female"  
 /tissue\_type="unfertilized egg"  
 /cell\_type="unfertilized egg"  
 /dev\_stage="unfertilized egg"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3-Pac; Site\_1: EcoRI; Site\_2: NotI; PolyA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. 'Normalization and subtraction: two approaches to facilitate gene discovery'. Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dm18 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT7T3-Pac vector. The library contained approximately 7.2 x 10<sup>5</sup> recombinants, with average insert sizes of 1-1.5 kb."  
 44 a 81 c 175 g 119 t

**BASE COUNT**  
**ORIGIN**

Query Match 87.0%; Score 17.4; DB 12; Length 419;  
 Best Local Similarity 94.7%; Pred. No. 4.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GCAGGAAAGGTTTGGTCTG 20  
 |||||  
 Db 42 GCTGAAAGGTTTGGTCTG 60

**RESULT 8**

**LOCUS**  
**DEFINITION**  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**

BE491715  
 db95a06.x1 Wellcome CRC psk animal cap Xenopus laevis cDNA clone  
 IMAGE:3380722 3', mRNA sequence.  
 BE491715  
 BE491715.1 GI:9611248  
 EST.  
 African clawed frog.  
 Xenopus laevis

**REFERENCE**  
**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

1 (bases 1 to 421)  
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R., Waterston, R., and Willson, R.  
 WashU Xenopus EST project, 1999  
 Unpublished (1999)  
 Contact: Sandy Clifton, Ph.D.  
 WashU Xenopus EST project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

Library constructed by N. Garrett, P. LeMaire, A.M. Zorn, and J.B. Gurdon (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 395.  
 Location/Qualifiers

## FEATURES

source

```
1. 421
  /organism="Xenopus laevis"
  /db_xref="taxon:8355"
  /clone="IMAGE:3380722"
  /clone_lib="Wellcome CRC PSK animal cap"
  /tissue_type="embryo, animal cap"
  /lab_host="DH10B (phage-resistant)"
  /note="Vector: pBluescript SK-; Site_1: NotI; Site_2:
  EcoRI; cDNAs were oligo-dT primed and directionally
  cloned. Staging according to Nieuwkoop and Faber. Library
  was constructed by N. Garrett, P. Lemaire, A.M. Zorn, and
  J.B. Gurdon (Wellcome/CRC Institute). Note: This is a
  Xenopus Gene Collection (XGC) library."
  BASE COUNT      49 a      82 c      171 g      119 t
  ORIGIN
```

Query Match 87.0%; Score 17.4; DB 10; Length 421;  
 Best Local Similarity 94.7%; Pred. No. 4.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY      2 GCAGGAAGGTTGGTCTG 20
      || ||||| ||||| |||||
  Db     24 GCTGGAAAGTTGGTCTG 42
```

RESULT 9  
 AW633486/c

LOCUS bl08d02.w1 Blackshear/Soares normalized Xenopus egg library Xenopus laevis cDNA clone PBX0008D02 5', mRNA sequence.  
 AW633486  
 AW633486.1 GI:7390567

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

African clawed frog.

Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 Xenopodinae; Xenopus.  
 1 (bases 1 to 431)

REFERENCE  
 AUTHORS Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.  
 Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman  
 , J.W., Bonaldo, M.F. and Soares, M.B.

TITLE The NIEHS Xenopus maternal EST project: interim analysis of the  
 first 13,879 ESTs from unfertilized eggs

JOURNAL Gene 267 (1), 71-87 (2001)  
 MEDLINE 21211403  
 COMMENT Contact: Perry J. Blackshear  
 Office of Clinical Research and Laboratory of Signal Transduction  
 National Institute of Environmental Health Sciences  
 A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,  
 USA

Tel: 919 541-4899  
 Fax: 919 541-4571

Email: black009@niehs.nih.gov

Clone is available through Research Genetics, Inc., 2130 Memorial  
 Parkway, Huntsville, AL 35901  
 phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email  
 cdna@resgen.com

DNA Sequencing and analyses performed by National Institutes of  
 Health Intramural Sequencing Center (NISC).

PCR Primers

FORWARD: TGTAACACGCGCAGT

BACKWARD: CAGGAACAGCTATGACC

Plate: 0008 row: D column: 02

Seq primer: T7 primer.

Location/Qualifiers

1. 431

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

## FEATURES

source

```
/clone="PBX0008D02"
/clone_lib="Blackshear/Soares normalized Xenopus egg
library"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH10B"
/note="Vector: pT73-Pac; Site_1: EcoRI; Site_2: NotI;
PolyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT73-Pac as described in Bonaldo, M.F., Lemmon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dT18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT73-Pac vector.
The library contained approximately 7.2 x 105
recombinants, with average insert sizes of 1-1.5 kb."
```

BASE COUNT 127 a 178 c 81 g 45 t  
 ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 431;  
 Best Local Similarity 94.7%; Pred. No. 4.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY      2 GCAGGAAGGTTGGTCTG 20
      || ||||| ||||| |||||
  Db     369 GCTGGAAAGTTGGTCTG 351
```

RESULT 10  
 AQ860919

LOCUS clone nbeb0016A15r, DNA sequence.  
 DEFINITION clone nbeb0016A15r, DNA sequence.  
 ACCESSION AQ860919  
 VERSION AQ860919.1 GI:6211376  
 KEYWORDS GSS.  
 SOURCE Oryza sativa.  
 ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.  
 1 (bases 1 to 434)  
 Wing, R.A. and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Rice Genome  
 Unpublished (1998)  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Seq primer: GGAACAGCTATGACCATG  
 Class: BAC ends  
 High quality sequence start: 23  
 High quality sequence stop: 337.  
 Location/Qualifiers

1. 434  
 /organism="Oryza sativa"  
 /strain="Japonica"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:4530"  
 /clone="nbeb0016A15r"  
 /clone\_lib="CUGI Rice BAC Library (EcoRI)"  
 /tissue\_type="Leaf"  
 /lab\_host="E. coli DH10B"  
 /note="Vector: pBACindigo; Site\_1: EcoRI; Site\_2: EcoRI;  
 Rice is the most important food crop in the world. Half of  
 the world population, especially those inhabiting highly

## FEATURES

source

populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center ([www.genome.clemson.edu](http://www.genome.clemson.edu))."

# BASE COUNT

90 a 109 c 116 g 118 t 1 others

Query Match 87.0%; Score 17.4; DB 17; Length 434;  
Best Local Similarity 94.7%; Pred. No. 4.1e+02; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCAGGAAGGTTGGTCT 19

Db 389 TCACAGAAAGTTGGTCT 407

# RESULT 11

AW633397/c

LOCUS ..

DEFINITION

AW633397

AW633397.1 GI:7390478

EST.

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 453)

Blackshear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.

Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman

The NIEHS Xenopus maternal EST project: interim analysis of the

Gene 267 (1), 71-87 (2001)

21211403

CONTACT: Perry J. Blackshear

Office of Clinical Research and Laboratory of Signal Transduction

National Institute of Environmental Health Sciences

A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709,

USA

Tel: 919 541-4899

Fax: 919 541-4571

Email: [black009@niehs.nih.gov](mailto:black009@niehs.nih.gov)

Clone is available through Research Genetics, Inc., 2130 Memorial

Parkway, Huntsville, AL 35901

phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email

[cdna@resgen.com](mailto:cdna@resgen.com)

DNA Sequencing and analyses performed by National Institutes of

Health Intramural Sequencing Center (NISC).

PCR Primers

FORWARD: TGTAAACGACGCCAGT

BACKWARD: CAGGAACAGCTATGACC

Plate: 0007 row: C column: 02

Seq primer: T7 primer.

Location/Qualifiers

1. .453

/organism="Xenopus laevis"

source

/db\_xref="taxon:8355"  
/clone="PBX0007C02"  
/clone\_lib="Blackshear/Soares normalized Xenopus egg  
library"

/sex="female"

/tissue\_type="unfertilized egg"

/cell\_type="unfertilized egg"

/dev\_stage="unfertilized egg"

/lab\_host="DH10B"

/notes="Vector: pT73-Pac; Site 1: EcoRI; Site 2: NotI;  
PolyA-selected mRNA was prepared from unfertilized Xenopus  
laevis eggs. The library was constructed in the vector  
pT73-Pac as described in Bonaldo, M.F., Lennon, G. and  
Soares, M.B. 'Normalization and subtraction: two  
approaches to facilitate gene discovery', Genome Research  
6:791-806, 1996. The first strand synthesis used a  
NotI-dT18 primer; double stranded cDNAs were ligated to  
EcoRI adapters, digested with NotI, and directionally  
cloned into the NotI and EcoRI-digested pT73-Pac vector.  
The library contained approximately 7.2 X 10<sup>5</sup>  
recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 133 a 182 c 48 t

# ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 453;

Best Local Similarity 94.7%; Pred. No. 4.2e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTGGTCTG 20

Db 391 GCTGGAAGGTTGGTCTG 373

# RESULT 12

BG016217/c

LOCUS ..

DEFINITION

IMAGE:3515801 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 453)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.

, Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person

, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

Waterston, R. and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

CONTACT: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

Library constructed by M. Kirschner (Harvard Medical School). DNA

Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.N.A.G.E. Consortium/LLNL at: [info@image.llnl.gov](mailto:info@image.llnl.gov)

Seq primer: -40RP from Gibco.

Location/Qualifiers

1. .455

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="IMAGE:3515801"

/clone\_lib="Kirschner embryo St10 14"

/tissue\_type="pooled embryos (stage 10-14)"

/lab\_host="DH10B (phage-resistant)"

# FEATURES

source

/note="vector: PCS2+; Site 1: NotI; Site 2: SalI;  
Size-selected for average insert size 1.2 kb. Library was  
constructed and donated by M. Kirschner (Harvard Medical  
School)."

BASE COUNT 126 a 177 c 97 g 55 t

## ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 455;  
Best Local Similarity 94.7%; Pred. No. 4.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTTGGTCTG 20

Db 423 GCTGGAAGGTTTGGTCTG 405

## RESULT 13

BG018230/c

LOCUS

DEFINITION

de66a08.y1 Kirschner embryo St10 14 Xenopus laevis cDNA clone

IMAGE:3516567 5', mRNA sequence.

ACCESSION BG018230

VERSION BG018230.1

KEYWORDS

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 463)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.

, Martin, J., Wyllie, T., Underwood, K., Theising, B., Bowers, Y., Person

, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

Waterston, R., and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by M. Kirschner (Harvard Medical School). DNA

Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -40RP from Gibco.

Location/Qualifiers

1..463

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="IMAGE:3516567"

/tissue\_type="pooled embryos (stage 10-14)"

/lab\_host="DH10B (phage-resistant)"

/note="vector: PCS2+; Site 1: NotI; Site 2: SalI;

Size-selected for average insert size 1.2 kb. Library was

constructed and donated by M. Kirschner (Harvard Medical

School)."

BASE COUNT 127 a 181 c 100 g 55 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 463;

Best Local Similarity 94.7%; Pred. No. 4.2e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTTGGTCTG 20

Db 426 GCTGGAAGGTTTGGTCTG 408

## RESULT 14

BG414006

LOCUS

DEFINITION

de76912.x1 Kirschner embryo St10 14 Xenopus laevis cDNA clone

IMAGE:3517655 3', mRNA sequence.

ACCESSION BG414006

VERSION BG414006.1

KEYWORDS

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 475)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.

, Martin, J., Wyllie, T., Underwood, K., Theising, B., Bowers, Y., Person

, B., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,

Waterston, R., and Wilson, R.

WashU Xenopus EST project, 1999

Unpublished (1999)

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by M. Kirschner (Harvard Medical School). DNA

Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clones from this library are available

through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov

Seq primer: -40RP from Gibco

High quality sequence stop: 460.

Location/Qualifiers

1..475

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/clone="IMAGE:3517655"

/tissue\_type="pooled embryos (stage 10-14)"

/lab\_host="DH10B (phage-resistant)"

/note="vector: PCS2+; Site 1: NotI; Site 2: SalI;

Size-selected for average insert size 1.2 kb. Library was

constructed and donated by M. Kirschner (Harvard Medical

School)."

BASE COUNT 57 a 98 c 185 g 135 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 475;

Best Local Similarity 94.7%; Pred. No. 4.2e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCAGGAAGGTTTGGTCTG 20

Db 60 GCTGGAAGGTTTGGTCTG 78

RESULT 15

BG264326/c

LOCUS

DEFINITION

de76912.y1 Kirschner embryo St10 14 Xenopus laevis cDNA clone

IMAGE:3517655 5', mRNA sequence.

ACCESSION BG264326

VERSION BG264326.1

KEYWORDS

SOURCE

ORGANISM

African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

1 (bases 1 to 488)

Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.

, Martin, J., Wyllie, T., Underwood, K., Theising, B., Bowers, Y., Person

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1980, Vol. 83, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1985, Vol. 10, No. 2, pp. 1-15
3. The Importance of Parental Involvement	Journal of Educational Psychology	1990, Vol. 82, No. 3, pp. 1-12
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1995, Vol. 98, No. 4, pp. 1-18
5. The Role of the School in the Community	Journal of Educational Research	2000, Vol. 103, No. 5, pp. 1-20
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2005, Vol. 108, No. 6, pp. 1-25
7. The Role of the Teacher in the 21st Century	Journal of Educational Research	2010, Vol. 113, No. 7, pp. 1-30
8. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2015, Vol. 118, No. 8, pp. 1-35
9. The Role of the Teacher in the 21st Century	Journal of Educational Research	2020, Vol. 123, No. 9, pp. 1-40
10. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2025, Vol. 128, No. 10, pp. 1-45

## FEATURES

BASE COUNT	140 a	186 c	104 g	58 t
ORIGIN				

```

Query Match      87.0%; Score 17.4; DB 12; Length 488;
Best Local Similarity 94.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Query      2 GCAGGAAAGGTTTGGTCTG 20
           ||| ||||| ||||| |||
DB          424 GCTGAAAGGTTTGGTCTG 406

```

Search completed: July 8, 2003, 09:21:51  
Job time : 1018.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 220.098 seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: us-09-647-780A-15  
Perfect score: 20  
Sequence: 1 gaacgcctcagagaagcctg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	20	100.0	20	6	AX014715	Sequence
2	20	100.0	2765	6	AX014701	Sequence
3	18.4	92.0	2583	10	AF302075	Mus muscu
4	18.4	92.0	2601	10	AF157106	Mus muscu
5	18.4	92.0	2652	10	AF302076	Mus muscu
6	18.4	92.0	2694	10	AF302077	Mus muscu
7	18.4	92.0	2892	10	AF157105	Mus muscu
8	18.4	92.0	2925	6	AX033272	Sequence
9	18.4	92.0	2925	10	AF176569	Mus muscu
10	18.4	92.0	3742	10	AF332142	Rattus no
11	18.4	92.0	65591	2	AC109980	Rattus no
12	18.4	92.0	114789	2	AC128969	Rattus no
13	18.4	92.0	151374	2	AC128214	Rattus no
14	18.4	92.0	189232	2	AC129646	Rattus no
15	18.4	92.0	188509	2	AC115933	Mus muscu
16	18.4	92.0	208249	2	AL607032	Mus muscu
17	18	90.0	83367	2	AC115654	Mus muscu
18	18	90.0	108873	2	AC095814	Rattus no
19	18	90.0	219409	2	AC117158	Rattus no
20	17.4	87.0	145550	9	AC092851	Homo sapi
21	17.4	87.0	146015	10	AL671877	Mouse DNA
22	17.4	87.0	171905	2	AC102074	Mus muscu
23	17.4	87.0	172871	10	AL607131	Mouse DNA
24	17.4	87.0	198459	2	AC079042	Mus muscu
25	17.4	87.0	233234	2	AL772311	Mus muscu
26	17.4	87.0	257586	2	AC127337	Mus muscu
27	17	85.0	64533	10	AY055726	Mus muscu
28	17	85.0	191918	2	AC102692	Mus muscu
29	17	85.0	210398	2	AC126041	Mus muscu
30	16.8	84.0	6842	10	AF014956	Rattus no
31	16.8	84.0	10840	1	AE000885	Xanthomonas
32	16.8	84.0	14190	1	AE000885	Methanoba
33	16.8	84.0	63749	9	AC004694	Homo sapi
34	16.8	84.0	68914	2	AC024970	Homo sapi
35	16.8	84.0	123302	2	AC095135	Rattus no
36	16.8	84.0	139227	2	AC120654	Rattus no
37	16.8	84.0	149602	2	AC095940	Rattus no
38	16.8	84.0	158079	2	AC122129	Homo sapi
39	16.8	84.0	159101	2	AC126205	Rattus no
40	16.8	84.0	160544	2	HS253P07	Homo sapi
41	16.8	84.0	161331	10	AL611950	Mouse DNA
42	16.8	84.0	168153	2	AL772289	Danio rer
43	16.8	84.0	170967	9	AC005023	Homo sapi
44	16.8	84.0	174992	2	AP005286	Oryza sat
45	16.8	84.0	176030	2	AC100735	Mus muscu

ALIGNMENTS

RESULT 1  
LOCUS AX014715  
DEFINITION Sequence 15 from Patent WO9953077.  
ACCESSION AX014715  
VERSION AX014715.1 GI:10040988  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE 1 (bases 1 to 20)  
AUTHORS Schwartz,J.C.; Gros,C.; Oulmet,T.; Rose,C.; Bonhomme,M.C. and  
Facchinetti,P.  
TITLE Novel pep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy

JOURNAL Patent: WO 9953077-A 15 21-OCT-1999;  
 INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
 CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
 MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

FEATURES  
 source  
 Location/Qualifiers  
 1..20  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="oligonucleotide"  
 6 a 6 c 6 g 2 t

BASE COUNT  
 ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGCCTCAGAGAGCCTG 20  
 |||||  
 Db 1 GAAGCCTCAGAGAGCCTG 20

RESULT 2  
 AX014701/c 2765 bp. DNA linear PAT 07-SEP-2000  
 LOCUS  
 DEFINITION Sequence 1 from Patent WO9953077.  
 AX014701  
 ACCESSION  
 VERSION AX014701.1 GI:10040975  
 KEYWORDS  
 SOURCE black rat.  
 ORGANISM Rattus rattus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 2765)  
 Schwartz,J.C., Gros,C., Oumet,T., Rose,C., Bonhomme,M.C. and  
 Facchinetti,P.  
 Novel nep ii membrane metalloprotease and its use for screening  
 inhibitors useful in therapy.  
 Patent: WO 9953077-A 1 21-OCT-1999;  
 INST-NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
 CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
 MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

FEATURES  
 source  
 Location/Qualifiers  
 1...2765  
 /organism="Rattus rattus"  
 /db\_xref="taxon:10117"  
 107..2431  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC07576.1"  
 /db\_xref="GI:10040976"  
 /translation="MGKSESSVGMERADNGRRRLGVEGGLVLLILLMGAIVTL  
 GVFSIGKQLPILNLSLHVRSHRTVVRVLRDSSQKSDICTTPSCVIAAARILQNM  
 QSKPKCNFYACGGWLRHVHVPETNSRYSVFDLRLEVLKGVLEDSVQRRA  
 VERAKTLRSCMNSQVIEKRDSEPLNLDIGGWPVAMKWNMTGPKWELERQAV  
 LNSQFNRLVLDLFWNDQNSRRVYIDOPTLGMPSREYFKEDSHRVEAYLOFM  
 TSVATRLRDNLPGCTDLVOEMAQVLLHLEHLANATVPROEKHDVTLVHRCLE  
 LQREFGLGFNFTLEIQNLVSVQVLELLPNEVYVYGIPILENLEEIDVDPQILQ  
 LVRLVLRIGLSQRFKARDYRKALYGTMEVWRRECVSVNSNMESAVGSLY  
 IKRAFSDKSLVSELIBEKIRSVFDNLNMDDESKKRAQKALNRIQIGYPDY  
 ILEDNRLHDEYSLSLFESEDLVFNGLQNLNNAORSLKRLKREKVDQNLWIGAAV  
 NAFYSNPNLIVFPAGILOPPFESKDOQALNFGGIGVIGHEITHGFDNCRGDKN  
 GMLDWSNFSARHFRQSQCMYQYNSFNELADNQNNGVSTLGENIADNGVGRQA  
 YKALQWLAEGRDQRLPGLNLTIAQLFFINVAQVWCYSRPFSAFQIKTDVHSPLK  
 YRVLSLQNLPGFSEAFHCPGRGSPMHPMNCRIW"  
 684 a 735 c 787 g 559 t

BASE COUNT  
 ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 8.6;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAGCCTCAGAGAGCCTG 20

Db 2377 GAAGCCTCAGAGAGCCTG 2358  
 |||||  
 RESULT 3  
 AF302075/c  
 LOCUS  
 DEFINITION 2583 bp mRNA linear ROD 11-JUN-2001  
 AF302075 Mus musculus neprilysin-like peptidase alpha mRNA, complete cds.  
 ACCESSION  
 VERSION AF302075.1 GI:10505359  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2583)  
 Shirotsani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,  
 Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,  
 Iwatsubo,T. and Saido,T.C.  
 Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most  
 rapidly and efficiently among thiorphan- and  
 phosphoramidon-sensitive endopeptidases  
 J. Biol. Chem. 276 (24), 21895-21901 (2001)  
 21293028  
 11278416  
 2 (bases 1 to 2583)  
 Shirotsani,K. and Saido,T.C.  
 Direct Submission  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

FEATURES  
 source  
 Location/Qualifiers  
 1..2583  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 25..2253  
 /note="endopeptidase"  
 /codon\_start=1  
 /product="neprilysin-like peptidase alpha"  
 /protein\_id="AAG18446.1"  
 /db\_xref="GI:10505360"  
 /translation="MVERAGWCKKSPGFVEYGLMVLILLLLGAIYVTLGVFYSIALRD  
 SLLKSDICTTPSCVIAAARILENMQSRNCFNYQYACGGWLRHVHVPETNSRYSVF  
 DILRDEVLKGVLEDSQHRPAVEKAKTLYRSCMNSQVIEKRDSEPLLSLVKMWG  
 GWPVAMKWNMTGKWELEKAVLEMTSVATLMDKQNLKSKESAMVREEMAEVLELET  
 LGMPSEYFQEDNNHKKVKEALEMTSVATLMDKQNLKSKESAMVREEMAEVLELET  
 HLANATVPROEKHDVTLVHRCLELQREFGLGFNFTLEIQNLVSVQVLELLPNEVYVYGIPILENLEEIDVDPQILQ  
 LVRLVLRIGLSQRFKARDYRKALYGTMEVWRRECVSVNSNMESAVGSLYIKRAFSDKSLVSELIBEKIRSVFDNLNMDDESKKRAQKALNRIQIGYPDY  
 ILEDNRLHDEYSLSLFESEDLVFNGLQNLNNAORSLKRLKREKVDQNLWIGAAV  
 NAFYSNPNLIVFPAGILOPPFESKDOQALNFGGIGVIGHEITHGFDNCRGDKN  
 GMLDWSNFSARHFRQSQCMYQYNSFNELADNQNNGVSTLGENIADNGVGRQA  
 YKALQWLAEGRDQRLPGLNLTIAQLFFINVAQVWCYSRPFSAFQIKTDVHSPLK  
 YRVLSLQNLPGFSEAFHCPGRGSPMHPMNCRIW"  
 665 a 667 c 736 g 515 t

BASE COUNT  
 ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 2583;  
 Best Local Similarity 95.0%; Pred. No. 63;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAAGCCTCAGAGAGCCTG 20  
 |||||  
 Db 2199 GAAGCCTCAGAGAGCCTG 2180

RESULT 4  
 AF157106/c  
 LOCUS  
 DEFINITION 2601 bp mRNA linear ROD 25-NOV-1999  
 AF157106 Mus musculus soluble secreted endopeptidase delta mRNA,  
 alternatively spliced product, complete cds.  
 ACCESSION  
 AF157106

```

VERSION      AF157106.1  GI:6467400
SOURCE       Mus musculus
ORGANISM     Mus musculus
REFERENCE    1 (bases 1 to 2601)
AUTHORS      Ikeda,K., Emoto,N., Raharjo,S.B., Nuhantari,Y., Saiki,K.,
              Yokoyama,M. and Matsuo,M.
TITLE       Molecular identification and characterization of novel
              membrane-bound metalloprotease, the soluble secreted form of which
              hydrolyzes a variety of vasoactive peptides
JOURNAL      J. Biol. Chem. 274 (45), 32469-32477 (1999)
MEDLINE      20011457
PUBMED       10542292
REFERENCE    2 (bases 1 to 2601)
AUTHORS      Ikeda,K., Emoto,N. and Matsuo,M.
TITLE       Direct Submission
JOURNAL      Submitted (08-JUN-1999) International Center for Medical Research,
              Kobe University School of Medicine, 7-5-1 Kusunoki, Chuo, Kobe
              6500017, Japan
FEATURES     Location/Qualifiers
              1..2601
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              59..2287
              /note="SEP(delta); metalloprotease; alternatively spliced"
              /codon_start=1
              /product="soluble secreted endopeptidase delta"
              /protein_id="AA13153.1"
              /db_xref="GI:6467401"
              /translation="MVERAGWCRKSPGVEYGLMVLVLLLLGALVTLGVFYSIGKQL
              SSLKSDICTPSCVTAARILEMDSQNPENFYQYACGWLRLHVPETNSRYSVF
              DILRDEVLKGVLEDSTQHRPAVERAKTLYRSCMNQSVTEKRDSEPLLSVLKMGV
              GWPVALDKWNETMGLKWELEQLAVNSQFNRRVLIDLFIMWDDQNSRHVYIDOPT
              HLNAVTPQEKRHVDYALYHRMDLMELOERFLGFNTLFIQNVLSVVEVLPDEE
              LTVVEYIPVLENDIIDSYSARTMONYLWRLVLDIGLSLRSORFKEARDYRKALYGT
              TVEEYWRCEVSYVNSMESAVGSLIIRKFSKSTVRELIEKIRSVFVNDLNLN
              WNDSEKKAEEKAMNIEQIGIPDYILEDDNNKHDEEYSSLTFFEDYFENGLQNLK
              FGAQSKLKLKREKVDQNLWIIIGAAVYNAFYSPNRNQIYFPAIGLQPPFFSKDQPSLN
              FGGTGMVIGHEITHGFDNDGRNFKNGMLDWSNFSARHFOQSQCMYQXGNSWE
              LADNONGVSGSLGNIADNGVROAYKAYLRWLADGGKDRGLGLNLTYAQLFNY
              AQVWCGSYRPEFAVQSIKTDVHSPKLYRVLSGLQNLPGFSEAFHCPGSPHMKRCR
              IW"
BASE COUNT   655 a 681 c 748 g 517 t
ORIGIN
Query Match      92.0%; Score 18.4; DB 10; Length 2601;
Best Local Similarity 95.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAAGCCTG 20
    ||| ||||| ||||| ||||| |||||
Db 2233 GAATGCCTCAGAGAAGCCTG 2214

RESULT 5
AF302076/c
LOCUS      AF302076
DEFINITION Mus musculus neprilysin-like peptidase beta mRNA, complete cds.
ACCESSION  AF302076
VERSION     AF302076.1  GI:10505361
KEYWORDS
SOURCE      Mus musculus
ORGANISM    Mus musculus
REFERENCE    1 (bases 1 to 2652)
AUTHORS      Shirohani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,
              Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,
              Iwatsubo,T. and Saido,T.C.
TITLE       Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most
              rapidly and efficiently among thiorphan- and
              phosphoramidon-sensitive endopeptidases
JOURNAL      J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE      11278416
PUBMED
REFERENCE    2 (bases 1 to 2694)
AUTHORS      Shirohani,K. and Saido,T.C.
TITLE       Direct Submission
JOURNAL      Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain
              Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama
              351-0198, Japan
FEATURES     Location/Qualifiers
              1..2652
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              25..2322
              /note="endopeptidase"
              /codon_start=1
              /product="neprilysin-like peptidase beta"
              /protein_id="AAG18447.1"
              /db_xref="GI:10505362"
              /translation="MVERAGWCRKSPGVEYGLMVLVLLLLGALVTLGVFYSIGKQL
              PLTSLHFSWDERTVYKRALRDSLSKSDICTPFCVIAAARILEMDSQNPENFY
              QYACGWLRLHVPETNSRYSVFDILRDEVLKGVLEDSTQHRPAVERAKTLYRS
              CMNQSVTEKRDSEPLLSVLKMGVGPVAMDKWNETMGLKWELEQLAVNSQFNRRVL
              IDLFIMWDDQNSRHVYIDOPTLGMPSREYFQEDNNKHVKRYALEFMTSVATMLRK
              DQNLKESAMVREEMAEVLETHLANATVPEQKRHDVTALYHRMDLMELOERFLGK
              FNTWLFIONVLSVVEVLPDEEVVYGYIPVLENDIIDSYSARTMONYLWRLVLD
              RIGLSLRSORFKEARDYRKALYGTVEEYWRCEVSYVNSMESAVGSLIIRKFSKDS
              KSTVRELIEKIRSVFVNDLNLNNDWDESKKAEKAMNIEQIGIPDYILEDDNNKH
              DEEYSSLTFFEDYFENGLQNLKNAQRSLKREKVDQNLWIIIGAAVYNAFYSPNRN
              QIYFPAIGLQPPFFSKDQPSLNFGTGMVIGHEITHGFDNDGRNFKNGMLDWSN
              FSAHFHFOQSQCMYQXGNSWELADNONGVSGSLGNIADNGVROAYKAYLRWL
              DGGKDRGLGLNLTYAQLFNYAQVWCGSYRPEFAVQSIKTDVHSPKLYRVLSGLQNL
              LPGFSEAFHCPGSPHMKRCIW"
BASE COUNT   682 a 685 c 755 g 530 t
ORIGIN
Query Match      92.0%; Score 18.4; DB 10; Length 2652;
Best Local Similarity 95.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAAGCCTG 20
    ||| ||||| ||||| ||||| |||||
Db 2268 GAATGCCTCAGAGAAGCCTG 2249

RESULT 6
AF302077/c
LOCUS      AF302077
DEFINITION Mus musculus neprilysin-like peptidase gamma mRNA, complete cds.
ACCESSION  AF302077
VERSION     AF302077.1  GI:10505363
KEYWORDS
SOURCE      Mus musculus
ORGANISM    Mus musculus
REFERENCE    1 (bases 1 to 2694)
AUTHORS      Shirohani,K., Tsubuki,S., Iwata,N., Takaki,Y., Harigaya,W.,
              Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,
              Iwatsubo,T. and Saido,T.C.
TITLE       Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most
              rapidly and efficiently among thiorphan- and
              phosphoramidon-sensitive endopeptidases
JOURNAL      J. Biol. Chem. 276 (24), 21895-21901 (2001)
MEDLINE      11278416
PUBMED
REFERENCE    2 (bases 1 to 2694)
AUTHORS      Shirohani,K. and Saido,T.C.
TITLE       Direct Submission
JOURNAL      Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain
              Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama
              351-0198, Japan

```

```
Query Match          92.0%; Score 18.4; DB 6; Length 2925;
Best Local Similarity 95.0%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 GAACGCCTCAGAGAAGCCTG 20
Db 2575 GAATGCTCAGAGAAGCCTG 2556

RESULT 9
AF176569/c
LOCUS
DEFINITION Mus musculus neprilysin-like metalloproteinase 1 (N11) mRNA,
complete cds.
ACCESSION AF176569
VERSION AF176569
KEYWORDS
SOURCE
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2925)
AUTHORS Ghadjar, G., Ruchon, A.F., DesGroselliers, L. and Boileau, G.
TITLE Molecular cloning and biochemical characterization of a new mouse
testis soluble-zinc-metalloproteinase of the neprilysin family
JOURNAL Biochem. J. 347 (Pt 2), 419-429 (2000)
MEDLINE 21060448
PUBMED 10749671
REFERENCE 2 (bases 1 to 2925)
AUTHORS Ghadjar, G., Ruchon, A.F., DesGroselliers, L. and Boileau, G.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-1999) Biochemistry, University of Montreal, 2900
Edouard Montpetit, Montreal, Que H3T 1J4, Canada
FEATURES
source
1. 2925
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="testis"
1. 2925
/gene="N11"
332. 2629
/gene="N11"
/feature="N11; metallo-endopeptidase"
/product="neprilysin-like metalloproteinase 1"
/protein_id="AAF69247.1"
/db_xref="GI:7769083"
/translation="MVERAGWCRKSPGVEYGLMVLVLLLLGAVITLVGVFYSIGKQL
PLTLLSFMDERTVVRKALDSKSDICTPSCVTAARILENMQDSRNPCEFY
QYACGWLHRIHVIPNTNRYSYDFDLDELEVLKGVEDSTSQHRAVPAKATLYRS
CMNQSVIKRDESPILLVLMKVGWPVAMDKWETMGLKWELOLAVLNSQFNRLV
IDLFTWDDQNSRHVIYIDPTLGMPSREYVFOEDNNHKKVKALEFMTSVATMLRK
DQNLKSAMVREEMAEVLETHLANATVPQEKRDHVTALYHRMDLMELQERGLKG
FNWTLFIONLVSSVEVLPDEEVVYIGIYLENLELIDISYARTMNYLVRLVLD
RIGLSQRFKARVDYRKALYGTVEEVVRECVSYVNSMESAVGSLYIKRAFSDS
KSTVRELIEKRSVFDNLDELNMDEESKKAQKAMNRIQIGYDPYIKRDNKHL
DEEYSLAFYEDLYFENGLOLKNNAQSLKLEKVDQNLWIIGAAVYNATYSPNRN
QIVFPAGILOPFFSKDQPSOLFSGIGVIGTHGFDGNGRDNKGMNDWSN
FSARHFOQSOCLMYQYGNFSEWELADQNVGFSILGENIADNGVGRQYKAYLRWLA
DGGKDRPLPLNLTYAQFFINIAQVWCGSYRPEFAVOSIKTVDHSPKLYRVLSGLQN
LPGFSEAPHCPRGSPMPMKRCRIW"
BASE COUNT 710 a 797 c 836 g 582 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 2925;
Best Local Similarity 95.0%; Pred. No. 62;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCCTCAGAGAAGCCTG 20
Db 2575 GAATGCTCAGAGAAGCCTG 2556

RESULT 10
AF332142/c
LOCUS
DEFINITION Rattus norvegicus chloride ion pump-associated 55 kDa protein
(Clp55) mRNA, complete cds.
ACCESSION AF332142
VERSION AF332142
KEYWORDS
SOURCE
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 3742)
AUTHORS Kitagawa, K., Yagyu, K., Yamamoto, A., Hattori, N., Omori, K., Zeng, X.T.
and Inagaki, C.
TITLE Molecular cloning and characterization of the Cl(-) pump-associated
55-kDa protein in rat brain
JOURNAL Biochem. Biophys. Res. Commun. 289 (2), 363-371 (2001)
MEDLINE 21573627
PUBMED 11716481
REFERENCE 2 (bases 1 to 3742)
AUTHORS Kitagawa, K., Yagyu, K., Hattori, N., Omori, K., Zeng, X.-T. and
Inagaki, C.
TITLE Direct Submission
JOURNAL Submitted (26-DEC-2000) Pharmacology, Kansai Medical University,
10-15 Fumizono-cho, Moriguchi, Osaka 570-8506, Japan
FEATURES
source
1. 3742
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/tissue_type="brain"
1. 3742
/gene="Clp55"
31. 1545
/gene="Clp55"
/codon_start=1
/product="chloride ion pump-associated 55 kDa protein"
/protein_id="AAK16548.1"
/db_xref="GI:13242075"
/translation="MGRFAATLVGSLFGLGLLLGLGLASAEPRAPKEKIAIVGAGI
GGTSSAYILRKFKGVKIDVFEREEIGGRATLKVGHDYEGAGSVTHPLNLMKRF
VKELGLSSYPASGLGVYNGKSLVPESSWFIINVLKVRYGFSQSLRHHWVEDLL
DFKRIYRQSHDYAFSSVEKLMIAIGDDIVRLNQLTRENKAKAGFSETFLNEMIA
PMKVNFGSTDLNAPFVAVSMATAADSNLWAVEGNKVCGLQASNSNLISVMS
IEKTRTKQGTGTPPKYEVYKTSSEIHSDFYDLVAAPLNKMSNITFRNFDPIEE
FNDPYQOLVTLIKGELNFTLFSRHKDQFGLSAILVTDSDMFINSLSIVASVSHKE
GPPAVDGMHVWTKFSKDIITKEQISKLFSDYAVRKPWLSYPYYPNPKPCPSILH
DRLYLINGIEFAASCMESAIAQYNALLAYHRWNGEDMIDQDDLYEKLKTEL"
BASE COUNT 989 a 833 c 905 g 1015 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 3742;
Best Local Similarity 95.0%; Pred. No. 61;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCCTCAGAGAAGCCTG 20
Db 666 GAATGCTCAGAGAAGCCTG 647

RESULT 11
AC109980
LOCUS
DEFINITION Rattus norvegicus clone CH230-31301, *** SEQUENCING IN PROGRESS
***, 31 unordered pieces.
ACCESSION AC109980
VERSION AC109980.3
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 65591)

```

## AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burtell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huiy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.D., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,K., Tang,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usman,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

Direct Submission  
Unpublished  
2 (bases 1 to 65591)  
Worley,K.C.

Direct Submission  
Submitted (09-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 65591)  
Worley,K.C.

Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:18860410.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: QGVV  
Center clone name: CH230-31301  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye: 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 27923 bases at least Q40  
Consensus quality: 30152 bases at least Q30  
Consensus quality: 32251 bases at least Q20  
-----

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently consists of 31 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1  
1114: contig of 1114 bp in length  
1214: gap of unknown length  
1215: contig of 1341 bp in length  
2555: gap of unknown length  
2556: contig of 1329 bp in length  
3984: gap of unknown length  
3985: contig of 1458 bp in length  
4085: gap of unknown length  
5542: gap of unknown length  
5543: contig of 1524 bp in length  
5643: gap of unknown length  
7166: gap of unknown length  
7266: contig of 1034 bp in length  
8300: gap of unknown length  
8401: contig of 1175 bp in length  
9575: gap of unknown length  
9576: contig of 1672 bp in length  
11347: gap of unknown length  
11348: contig of 1581 bp in length  
13028: gap of unknown length  
13029: contig of 1746 bp in length  
13129: gap of unknown length  
14875: contig of 1733 bp in length  
14876: gap of unknown length  
16807: gap of unknown length  
16808: contig of 1167 bp in length  
17974: gap of unknown length  
17975: contig of 1520 bp in length  
18075: gap of unknown length  
19594: gap of unknown length  
21242: contig of 1548 bp in length  
21243: gap of unknown length  
23056: contig of 1714 bp in length  
23156: gap of unknown length  
23157: contig of 1510 bp in length  
24667: gap of unknown length  
24668: contig of 1642 bp in length  
26409: gap of unknown length  
26509: contig of 2158 bp in length  
28666: gap of unknown length  
28667: contig of 1370 bp in length  
30136: gap of unknown length  
30237: contig of 1552 bp in length  
31788: gap of unknown length  
31889: contig of 2285 bp in length  
34173: gap of unknown length  
34174: contig of 1686 bp in length  
34274: gap of unknown length  
35959: gap of unknown length  
36059: contig of 2197 bp in length  
38256: gap of unknown length  
38357: contig of 3780 bp in length  
42136: gap of unknown length  
42137: contig of 2035 bp in length  
44237: gap of unknown length  
44371: contig of 2562 bp in length  
46933: gap of unknown length  
46934: contig of 2027 bp in length  
47034: gap of unknown length  
49061: contig of 2703 bp in length  
49161: gap of unknown length  
51864: contig of 4805 bp in length  
51964: gap of unknown length  
56769: contig of 2856 bp in length  
56869: gap of unknown length  
59725: contig of 5767 bp in length  
59825: gap of unknown length  
65591: contig of 5767 bp in length.

## FEATURES

Location/Qualifiers  
1..65591  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"



34616 34715: gap of unknown length  
 36615: contig of 1900 bp in length  
 36715: gap of unknown length  
 38473: contig of 1758 bp in length  
 38573: gap of unknown length  
 38574 40605: contig of 2032 bp in length  
 40705: gap of unknown length  
 42853: contig of 2148 bp in length  
 42953: gap of unknown length  
 45191: contig of 2238 bp in length  
 45291: gap of unknown length  
 47147: contig of 1856 bp in length  
 47247: gap of unknown length  
 47248 48910: contig of 1663 bp in length  
 48910: gap of unknown length  
 50703: contig of 1693 bp in length  
 50803: gap of unknown length  
 52366: contig of 1563 bp in length  
 52466: gap of unknown length  
 52467 54624: contig of 2158 bp in length  
 54625 54724: gap of unknown length  
 54725 57204: contig of 2480 bp in length  
 57205 57304: gap of unknown length  
 57305 59915: contig of 2611 bp in length  
 59916 60015: gap of unknown length  
 60016 63063: contig of 3048 bp in length  
 63064 63163: gap of unknown length  
 63164 65786: contig of 2623 bp in length  
 65787 65886: gap of unknown length  
 65887 68161: contig of 2275 bp in length  
 68162 68261: gap of unknown length  
 68262 71290: contig of 3029 bp in length  
 71291 71390: gap of unknown length  
 71391 73552: contig of 2162 bp in length  
 73553 73652: gap of unknown length  
 73653 75936: contig of 2284 bp in length  
 75937 76037: gap of unknown length  
 76037 79206: contig of 3170 bp in length  
 79207 79306: gap of unknown length  
 79307 82322: contig of 3016 bp in length  
 82323 82422: gap of unknown length  
 82423 86777: contig of 4355 bp in length  
 86778 86877: gap of unknown length  
 86878 90387: contig of 3510 bp in length  
 90388 90487: gap of unknown length  
 90488 94881: contig of 4394 bp in length  
 94882 94981: gap of unknown length  
 94982 99220: contig of 4239 bp in length  
 99221 99320: gap of unknown length  
 99321 104322: contig of 4902 bp in length  
 104323 104322: gap of unknown length  
 104323 109517: contig of 5195 bp in length  
 109518 109617: gap of unknown length  
 109618 114789: contig of 5172 bp in length.

## FEATURES.

source

Location/Qualifiers

1..114789

/organism="Rattus norvegicus"

/db\_xref="taxon:10116"

/clone="CH230-283G2"

BASE COUNT 30903 a 22048 c 22499 g 31310 t 8029 others

## ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 114789;  
 Best Local Similarity 95.0%; Pred. No. 45;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCCCTCAGAGAGCCTG 20  
 ||||| ||||| ||||| ||||| |||||

Db 50624 GAACACTTCAGAGAGCCTG 50643

RESULT 13  
 AC128214

LOCUS AC128214 151374 bp DNA linear HTG 19-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-474C23, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 36 unordered pieces.  
 AC128214  
 AC128214.1 GI:21908823  
 HTG: HTGS\_PHASE1.  
 KEYWORDS Rattus norvegicus.  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 151374)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaralungue,H.C., Are,J.R., Ayelle,M., Banks,T.,  
 Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,J., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
 Homsli,P., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Jacobson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,  
 Oregunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,Y.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 151374)  
 Worley,K.C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: G2MG  
 Center clone name: CH230-474C23  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 111011 bases at least Q40

Consensus quality: 115816 bases at least Q30  
Consensus quality: 119457 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length.  
(see [http://www.ngsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.ngsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 36 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1302: contig of 1302 bp in length  
1303 1402: gap of unknown length  
1403 2544: contig of 1142 bp in length  
2545 2644: gap of unknown length  
2645 4286: contig of 1642 bp in length  
4287 4386: gap of unknown length  
4387 5519: contig of 1133 bp in length  
5520 5619: gap of unknown length  
5620 6667: contig of 1048 bp in length  
6668 8113: gap of unknown length  
8114 8213: gap of unknown length  
8214 10020: contig of 1807 bp in length  
10021 10120: gap of unknown length  
10121 11659: contig of 1539 bp in length  
11660 11759: gap of unknown length  
11760 13156: contig of 1397 bp in length  
13157 13256: gap of unknown length  
13257 14506: contig of 1250 bp in length  
14507 14606: gap of unknown length  
14608 16593: contig of 1987 bp in length  
16594 16693: gap of unknown length  
16694 18251: contig of 1558 bp in length  
18252 20172: gap of unknown length  
20173 20272: gap of unknown length  
20273 21327: contig of 1055 bp in length  
21328 21427: gap of unknown length  
21429 23523: contig of 2096 bp in length  
23524 23623: gap of unknown length  
23624 25982: contig of 2359 bp in length  
25983 26083: gap of unknown length  
26084 27163: contig of 1081 bp in length  
27164 29094: gap of unknown length  
29095 29194: gap of unknown length  
29195 31569: contig of 2375 bp in length  
31570 31669: gap of unknown length  
31670 35290: contig of 3621 bp in length  
35291 35390: gap of unknown length  
35391 38443: contig of 3053 bp in length  
38444 41858: gap of unknown length  
41859 41958: contig of 3315 bp in length  
41959 44884: gap of unknown length  
44885 44984: gap of unknown length  
44985 50497: contig of 5513 bp in length  
50498 50597: gap of unknown length  
50599 53825: contig of 3228 bp in length  
53826 53925: gap of unknown length  
53926 57726: contig of 3801 bp in length  
57727 57826: gap of unknown length  
57827 63827: contig of 5901 bp in length  
63828 70302: contig of 6475 bp in length  
70303 70402: gap of unknown length  
70403 75436: contig of 5034 bp in length  
75437 81613: contig of 6077 bp in length  
81614 81713: gap of unknown length

\* 81714 90032: contig of 8319 bp in length  
90033 90132: gap of unknown length  
90133 99154: contig of 9022 bp in length  
99155 99254: gap of unknown length  
99255 110498: contig of 11244 bp in length  
110499 110598: gap of unknown length  
110599 121332: contig of 10734 bp in length  
121333 121432: gap of unknown length  
121433 133140: contig of 11708 bp in length  
133141 133240: gap of unknown length  
133241 151374: contig of 18134 bp in length.

FEATURES  
1..151374 Location/Qualifiers

Source

/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-474C23"

BASE COUNT 41445 a 31347 c 30824 g 42504 t 5254 others  
ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 151374;  
Best Local Similarity 95.0%; Pred. NO. 44;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAAGCCCTCAGAGAGCCTG 20  
|||||

Db 43420 GAATGCCTCAGAGAGCCTG 43439

RESULT 14

AC129646

LOCUS AC129646 169232 bp DNA linear HTG 31-JUL-2002  
DEFINITION Rattus norvegicus clone CH230-27G11, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 74 unordered pieces.

ACCESSION AC129646.1 GI:22024414

VERSION HTG; HTGS\_PHASE1

KEYWORDS Norway rat.

SOURCE Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus

REFERENCE

AUTHORS

1 (bases 1 to 169232)  
Muzny D.M., Adams C., Adio-Oduola B., Ali-Osman F.R., Allen C.,  
Alsbrooks S.L., Amaralunge H.C., Are J.R., Ayale M., Banks T.,  
Barbaria J., Benton J., Blinage K., Blankenburg K., Bonnin D.,  
Bouck J., Bowie S., Brieva M., Brown E., Brown M., Bryant N.P.,  
Buhay C., Burch P., Burkett C., Burrell K.L., Byrd N.C.,  
Caron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,  
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,  
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,  
Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,  
Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinah H.H.,  
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,  
Einhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,  
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,  
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,  
Gorrell J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,  
Harris C., Harris K., Hart M., Havlak P., Hawes A., Hernandez J.,  
Hernandez O., Hodgson A., Hognes M., Holloway C., Hollins B.,  
Honsi F., Howard S., Huber J., Huiy K.S., Hume J., Jackson L.E.,  
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,  
Karlsson E., Kelly S., Khan U., King L., Korvah J., Kovar C.,  
Kratochvil J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,  
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulseged H.,  
Lozano R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,  
Maheshwari M., Mapua P., Martin R., Martindale A., Martinez E.,  
Massey E., Mawhiney E., McLeod M.P., Meador M., Mei G., Metzker M.,  
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,  
Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,  
Nguyen N., Nickerson E., Nwokenwo S., Ogih M., Okunuga G.,  
Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,  
Peters L., Pickens R., Primus E., Pu L.L., Quiles M., Ren Y.,  
Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savary G.,



Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 188509)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,  
Chazarar, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
Ezaro, S., Ferreira, P., Fitzgerald, M., FitzHugh, W., Gage, D.,  
Galgagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,  
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRocque, K.,

```
misc_feature
misc_feature
misc_feature
misc_feature
```

* 1	1958:	contig of 1998 bp in length
* 1959	2058:	gap of 100 bp
* 2059	17053:	contig of 14955 bp in length
* 17054	17153:	gap of 100 bp
* 17154	38916:	contig of 21763 bp in length
* 38917	39016:	gap of 100 bp
* 39017	78433:	contig of 39417 bp in length
* 78434	78533:	gap of 100 bp
* 78534	116274:	contig of 37741 bp in length
* 116275	116374:	gap of 100 bp
* 116375	189509:	contig of 72135 bp in length

Location/Qualifiers

1. 198509  
/organization="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP24-53918"  
/clone\_lbb="RPCI-24 Male  
1. 1958  
/notes="assembly\_fragment"  
2059. 17053  
/note="assembly\_fragment"  
17154. 38916  
/notes="assembly\_fragment"  
39017. 78433

misc\_feature /note="assembly\_fragment"  
78534.116274  
/note="assembly\_fragment"  
misc\_feature 116375.188509  
/note="assembly\_fragment"  
BASE COUNT 58636 a 36706 c 36824 g 55842 t 501 others  
ORIGIN

Query Match 92.0%; Score 18.4; DB 2; Length 188509;  
Best Local Similarity 95.0%; Pred. No. 43;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCCTCAGAGCCTG 20  
||||||| |||||||||  
Db 175004 GAACGCCTCAGAGCCTG 174985

Search completed: July 8, 2003, 03:34:57  
Job time : 224.098 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 125.659 seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-15

Perfect score: 20

Sequence: 1 gaacgcctcagagaagcctg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %		Length	DB ID	Description
		Match				
1	20	100.0	20	AAZ28821		Rat membrane metal
c	20	100.0	2286	24	AAD28130	Soluble secreted e
c	20	100.0	2765	20	AAZ28810	Rat membrane metal
c	18.4	92.0	2925	21	AAA63763	CDNA encoding neut
5	16.4	82.0	1836	17	AAZ29734	Brevibacterium fla
6	16.4	82.0	1839	22	AAH66364	C glutamicum codin
7	16.4	82.0	2952	21	AAA71996	C glutamicum ilvd
8	16.4	82.0	2952	21	AAA40283	C glutamicum ilvd
c	16.4	82.0	34980	22	AAH68528	C glutamicum codin

C 10	16	80.0	2480	22	AAH99656	Human protein enco
C 11	15.8	79.0	155	21	AAC20810	Human secreted pro
C 12	15.8	79.0	267	24	AA598541	Human protective D
C 13	15.8	79.0	274	21	AAF11088	Fusarium venenatum
C 14	15.8	79.0	321	24	AA598540	Human protective D
C 15	15.8	79.0	373	16	AAT19761	Human gene signatu
C 16	15.8	79.0	401	22	AAK96038	Human gene regulin g
C 17	15.8	79.0	401	22	AAK97531	Human neurergulin g
C 18	15.8	79.0	442	22	ABA45455	Human breast cell
C 19	15.8	79.0	442	22	ABA55953	Human foetal liver
C 20	15.8	79.0	442	22	ABA25622	Probe #4088 for ge
C 21	15.8	79.0	442	22	AAK04167	Human brain expres
C 22	15.8	79.0	442	22	AAK29651	Human bone marrow
C 23	15.8	79.0	442	22	AAI14227	Probe #4160 for ge
C 24	15.8	79.0	442	22	AAI35609	Probe #4295 used t
C 25	15.8	79.0	442	22	AAI04064	Probe #4055 used t
C 26	15.8	79.0	442	24	ABS04205	Human genome-deriv
C 27	15.8	79.0	453	24	ABK38790	CDNA encoding lung
C 28	15.8	79.0	478	24	ABN88230	Human colon cancer
C 29	15.8	79.0	495	24	ABL38113	Human colon tumour
C 30	15.8	79.0	616	24	ABO58559	Human colon cancer
C 31	15.8	79.0	652	21	AAZ80421	Human colon cancer
C 32	15.8	79.0	934	22	AAK61467	Human immune/haema
C 33	15.8	79.0	981	22	AAK80202	Human immune/haema
C 34	15.8	79.0	1067	22	AAK74763	Human immune/haema
C 35	15.8	79.0	1077	22	AAI13363	Human secreted pro
C 36	15.8	79.0	1230	21	AAAC0785	Arabidopsis thalia
C 37	15.8	79.0	1738	21	AAAC45384	Arabidopsis thalia
C 38	15.8	79.0	1741	21	AAAC40906	Arabidopsis thalia
C 39	15.8	79.0	1815	14	AAQ42188	ODC CDNA. Synthet
C 40	15.8	79.0	1851	21	AAAC47762	zea mays DNA fragm
C 41	15.8	79.0	1958	21	AAF18267	Lung cancer associ
C 42	15.8	79.0	1959	22	AAH33298	Human colon cancer
C 43	15.8	79.0	1978	22	AAAS03902	Human secreted pro
C 44	15.8	79.0	2377	24	AA598520	Human protective D
C 45	15.8	79.0	2395	23	ABV27808	Human prostate exp

#### ALIGNMENTS

RESULT 1  
AAZ28821  
ID AAZ28821 standard; DNA; 20 BP.  
AC AAZ28821;  
XX  
XX  
DT 01-FEB-2000 (first entry)  
DE Rat membrane metalloprotease NEPII gene probe #11.  
XX  
XX  
KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation; ss;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.  
XX  
XX  
OS Synthetic.  
OS Rattus rattus.  
XX  
XX  
PN FR277291-A1.  
XX  
PD 15-OCT-1999.  
XX  
XX  
PF 08-APR-1998; 98FR-0004389.  
XX  
XX  
PR 08-APR-1998; 98FR-0004389.  
XX  
XX  
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX  
PI Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
XX  
XX  
DR WPI; 1999-593429/51.

XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease -  
XX  
PS Claim 3; Page 22; 29pp; French.  
XX  
CC Sequences AA228811-228827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysine II (NEPII) gene  
CC (AA228810). NEPII is involved in (in)activation of neuronal and  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
CC disorders.  
XX  
SQ Sequence 20 BP; 6 A; 6 C; 6 G; 2 T; 0 other;  
  
Query Match 100.0%; Score 20; DB 20; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GAAGCGCTCAGAGAGCGCTG 20  
DB 1 GAAGCGCTCAGAGAGCGCTG 20  
  
RESULT 2  
AA228130/c  
ID AAD28130 standard; DNA; 2286 BP.  
XX  
AC AAD28130;  
XX  
DT 07-MAY-2002 (first entry)  
XX  
DE Soluble secreted endopeptidase (SEP) consensus DNA.  
XX  
KW Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
KW FSAD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
XX  
OS Homo sapiens.  
OS Mus sp.  
OS Rattus sp.  
XX  
XX Key Location/Qualifiers  
FH misc\_feature 1664..2286  
FT /\*tag= a  
FT /note= "Encodes catalytic domain"  
XX  
XX WO200206492-A1.  
XX  
XX 24-JAN-2002.  
XX  
XX 16-JUL-2001; 2001WO-IB01363.  
XX  
XX 14-JUL-2000; 2000GB-0017387.  
XX  
XX (PFIZ ) PFIZER LTD.  
XX (PFIZ ) PFIZER INC.  
XX  
XX Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
XX WPI; 2002-155042/20.  
XX  
XX An isolated and/or purified nucleic acid encoding a human soluble  
PT secreted endopeptidase which is useful for treating sexual dysfunction,  
PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
PT female sexual arousal disorder -  
XX

PS Disclosure; Fig 6; 167pp; English.  
XX  
CC The invention relates to an isolated and/or purified nucleic acid  
CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
CC antibody and the compound which inhibits or selectively inhibits the  
CC human SEP protein are useful in the manufacture of a medicament for  
CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
CC also useful for treating the above disorders and other disorders such  
CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
CC hypoaactive sexual desire disorder. The present sequence is SEP consensus  
CC DNA sequence found in human, mouse and rat.  
XX  
SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
  
Query Match 100.0%; Score 20; DB 24; Length 2286;  
Best Local Similarity 100.0%; Pred. No. 2.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 GAAGCGCTCAGAGAGCGCTG 20  
DB 2235 GAAGCGCTCAGAGAGCGCTG 2216  
  
RESULT 3  
AAZ28810/c  
ID AAZ28810 standard; cDNA; 2765 BP.  
XX  
AC AAZ28810;  
XX  
DT 01-FEB-2000 (first entry)  
XX  
DE Rat membrane metalloprotease NEPII gene.  
XX  
KW Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
XX  
OS Rattus rattus.  
XX  
XX FR2777291-A1.  
XX  
XX 15-OCT-1999.  
XX  
XX 08-APR-1998; 98FR-0004389.  
XX  
XX 08-APR-1998; 98FR-0004389.  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX Ouimet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
XX Schwartz JC;  
XX  
XX WPI; 1999-593429/51.  
XX P-PSDB; AAY44177.  
XX  
XX New membrane metalloprotease NEP II, involved in proteolysis of  
XX neuronal and hormonal peptides, used to screen for inhibitors,  
XX potentially useful for treating e.g. cardiovascular disease -  
XX  
XX Claim 2; Page 12-16; 29pp; French.  
XX  
CC This sequence represents the gene for the rat membrane metalloprotease  
CC designated neprilysine II (NEPII), which is involved in (in)activation  
CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
CC for specific substrates (used to detect NEPII in cells and tissues) or  
CC inhibitors, which can also be used to detect NEPII or for treatment of  
CC disorders related to peptidergic signalling in which NEPII is involved,  
CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
CC endocrine disorders.

XX SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
 Query Match 100.0%; Score 20; DB 20; Length 2765;  
 Best Local Similarity 100.0%; Pred. No. 2.4; Length 2765;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAAGCCCTCAGAGAGCCTG 20  
 |||||  
 Db 2377 GAAGCCCTCAGAGAGCCTG 2358

RESULT 4  
 AAA63763/c  
 ID AAA63763 standard; cDNA; 2925 BP.  
 XX AC  
 XX AA63763;  
 XX 04-DEC-2000 (first entry)  
 XX cDNA encoding neutral endopeptidase metalloproteinase-like enzyme NL-1.  
 XX Neprilysin; neutral endopeptidase metalloproteinase-like enzyme;  
 KW NEP-like enzyme; protein production; protein secretion;  
 KW neurological disease; Alzheimer's disease; pain; psychiatric disorder;  
 KW fertility; bone disease; abnormal phosphate metabolism; ss.  
 XX OS Mus sp.  
 XX FH Key Location/Qualifiers  
 XX CDS 332..2629  
 XX FT /\*tag= a  
 XX FT /product= "neutral endopeptidase metalloproteinase-like  
 XX enzyme NL-1"  
 XX PN  
 XX WO2000047750-A2.  
 XX 17-AUG-2000.  
 XX 11-FEB-2000; 2000WO-CA00147.  
 XX 11-FEB-1999; 99CA-2260376.  
 XX (UYMO-) UNIV MONTREAL.  
 XX Desgroselliers L, Boileau G;  
 XX WPI; 2000-549148/50.  
 XX P-PSDB; AAB08130.  
 XX Novel neutral endopeptidase-like metalloproteinase polypeptides and  
 FT polynucleotides, used to screen for related sequences and enzyme  
 FT inhibitors, used for the treatment of NL-3 related bone disorders -  
 XX Disclosure; Fig 3; 59pp; English.  
 XX PS  
 XX The present sequence encodes a murine neutral endopeptidase  
 CC metalloproteinase-like enzyme, designated NL-1. The specification  
 CC also describes NL-2 and NL-3. The NL enzymes are used to test for  
 CC specific inhibitors. The N-terminal region of the enzymes can be used  
 CC to promote production and secretion of foreign proteins and active  
 CC biopeptides, using chimeric constructs containing the foreign protein  
 CC downstream from and in phase with the N-terminal region. The NL enzymes  
 CC are have been localised to the brain, and may be useful in the  
 CC treatment of neurological diseases such as Alzheimer's disease, pain,  
 CC and psychiatric disorders. NL enzymes have also been localised to the  
 CC testis and ovaries, and may be used to control fertility. They have  
 CC also been localised to bones, and may be used to treat bone diseases,  
 CC and abnormal phosphate metabolisms related to improper peptide  
 CC processing by the NL-3 enzyme.  
 XX SQ Sequence 2925 BP; 710 A; 797 C; 836 G; 582 T; 0 other;

Query Match 92.0%; Score 18.4; DB 21; Length 2925;  
 Best Local Similarity 95.0%; Pred. No. 15;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAAGCCCTCAGAGAGCCTG 20  
 |||||  
 Db 2575 GAATGCTCAGAGAGCCTG 2556

RESULT 5  
 AAT29734  
 ID AAT29734 standard; DNA; 1836 BP.  
 XX AC  
 XX AAT29734;  
 XX 27-NOV-1996 (first entry)  
 XX Brevibacterium flavum MJ-233 dihydroxy acid dehydratase coding region.  
 XX Dihydroxy-acid dehydratase; EC 4.2.1.9; amino acid biosynthesis;  
 KW isoleucine; Ile; valine; Val; Coryneform bacterium; ds.  
 XX OS Brevibacterium flavum MJ-233.  
 XX FH Key Location/Qualifiers  
 XX mat\_peptide 1..1836  
 XX FT /\*tag= a  
 XX FT /EC\_number= 4.2.1.9  
 XX PN JP08089249-A.  
 XX 09-APR-1996.  
 XX 29-SEP-1994; 94JP-0234612.  
 XX 29-SEP-1994; 94JP-0234612.  
 XX (MITU) MITSUBISHI CHEM CORP.  
 XX WPI; 1996-233342/24.  
 XX P-PSDB; AAR91947.  
 XX DNA fragment contg. a gene coding for dihydroxy-acid dehydratase -  
 PT derived from Brevibacterium flavum MJ 233, useful for prodn. of  
 PT L-isoleucine and L-valine  
 XX PS Claim 5; Page 9-11; 12pp; Japanese.  
 XX A 6 kb genomic DNA KpnI fragment containing an open reading frame  
 CC (ORF) coding for dihydroxy-acid dehydratase was isolated from  
 CC Brevibacterium flavum MJ-233. The sequence of the 1836 bp ORF was  
 CC determined (i.e. the present sequence) and was found to encode a  
 CC protein of 612 amino acids. The enzyme is involved in biosynthesis  
 CC of the amino acids isoleucine and valine.  
 XX SQ Sequence 1836 BP; 426 A; 606 C; 475 G; 329 T; 0 other;

Query Match 82.0%; Score 16.4; DB 17; Length 1836;  
 Best Local Similarity 94.4%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GAAGCCCTCAGAGAGCCTG 18  
 |||||  
 Db 1719 GAAGCCCTCAGAGAGCCTG 1736

RESULT 6  
 AAH66364  
 ID AAH66364 standard; DNA; 1839 BP.  
 XX AC AAH66364;  
 XX 26-SEP-2001 (first entry)

xx DE C glutamicum coding sequence fragment SEQ ID NO: 1399.  
 xx KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 xx KW organic acid synthesis; ds.  
 xx OS Corynebacterium glutamicum.  
 xx PN EPI108790-A2.  
 xx PD 20-JUN-2001.  
 xx PF 18-DEC-2000; 2000EP-0127688.  
 xx PR 16-DEC-1999; 99JP-0377484.  
 xx PR 07-APR-2000; 2000JP-0159162.  
 xx PR 03-AUG-2000; 2000JP-0280988.  
 xx PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 xx PI Nakagawa S, Mizoquchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 xx PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 xx DR WPI; 2001-376931/40.  
 xx DR P-PSDB; AAG91145.  
 xx PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 xx PT mutation point of a gene, measuring expression of a gene, analysing  
 xx PT expression profile or pattern of a gene and identifying homologous gene  
 xx PS Claim 8; SEQ ID NO: 1399; 246pp + Sequence Listing; English.  
 xx CC The present invention provides a number of nucleotide and protein  
 xx CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 xx CC are useful for identifying the mutation point of a gene derived from a  
 xx CC mutant of coryneform bacterium, measuring expression amount and  
 xx CC analysing the expression profile or expression pattern of a gene derived  
 xx CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 xx CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 xx CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 xx CC particularly L-lysine. The present sequence is a nucleic acid described  
 xx CC in the exemplification of the invention.  
 xx CC Note: The sequence data for this patent did not form part of the printed  
 xx CC specification, but was obtained in electronic format directly from the  
 xx CC European Patent Office.  
 xx SQ Sequence 1839 BP; 423 A; 610 C; 478 G; 328 T; 0 other;  
 Query Match 82.0%; Score 16.4; DB 22; Length 1839;  
 Best Local Similarity 94.4%; Pred. No. 1.4e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GAAGCGCTCAGAGAAGCC 18  
 Db 1722 GAAGCGCTCCGAGAAGCC 1739  
 RESULT 7  
 AAA71996  
 ID AAA71996 standard; DNA; 2952 BP.  
 AC AAA71996;  
 XX 19-JAN-2001 (first entry)  
 DE C. glutamicum ilvD DNA.  
 KW D-pantothenate biosynthesis; panBC; ilvA; L-valine production; ilvD;  
 KW dihydroxyacid dehydratase; ilvD; acetoxyhydroxy acid synthase; ilvBN;  
 KW isomeroreductase; ilvC; nutrition; threonine dehydratase; ds.  
 OS Corynebacterium glutamicum.

xx DE19907567-A1.  
 xx PD 24-AUG-2000.  
 xx PF 22-FEB-1999; 99DE-1007567.  
 xx PR 22-FEB-1999; 99DE-1007567.  
 xx PA (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
 xx PI Eggeling L, Sahn H;  
 xx DR WPI; 2000-566215/53.  
 xx DR P-PSDB; AAB10681.  
 xx PT Microbial production of L-valine useful as medicine and in nutrition  
 xx PT comprises growing bacteria that have increased activity of selected  
 xx PT genes, e.g., dihydroxyacid dehydratase  
 xx Example 1; Page 10-12; 24pp; German.  
 xx CC This invention describes a novel method for the microbial production of  
 xx CC L-valine (I) which uses a microorganism in which either the dihydroxyacid  
 xx CC dehydratase (ilvD) activity and/or gene expression or the acetoxyhydroxy  
 xx CC acid synthase (ilvBN) and isomeroreductase (ilvC) activity and/or gene  
 xx CC expression has been increased (I) is useful in animal and human  
 xx CC nutrition and as a medicine. Increasing expression of the dihydroxyacid  
 xx CC dehydratase, BN and/or C genes results in increased yields of (I),  
 xx CC particularly when used in conjunction with inactivation of genes involved  
 xx CC in synthesis of D-pantothenate e.g. threonine dehydratase (ilvA). This  
 xx CC sequence encodes the Corynebacterium glutamicum ilvD protein which is  
 xx CC used to illustrate the method of the invention.  
 xx SQ Sequence 2952 BP; 710 A; 933 C; 743 G; 566 T; 0 other;  
 Query Match 82.0%; Score 16.4; DB 21; Length 2952;  
 Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GAAGCGCTCAGAGAAGCC 18  
 Db 2008 GAAGCGCTCCGAGAAGCC 2025  
 RESULT 8  
 AAA40283  
 ID AAA40283 standard; DNA; 2952 BP.  
 AC AAA40283;  
 XX 02-NOV-2000 (first entry)  
 DE C. glutamicum ilvD genomic DNA.  
 KW D-pantothenic acid; panB; panC; ilvD; pantotheanate synthetase;  
 KW ketopantoatehydroxymethyltransferase; dihydroxyaciddehydratase;  
 KW vitamin; ds.  
 OS Corynebacterium glutamicum.  
 FH Key Location/Qualifiers  
 FT CDS 290..2128  
 FT /\*tag= a  
 FT /product= "ilvD"  
 XX EPI006189-A2.  
 xx PD 07-JUN-2000.  
 xx PF 30-NOV-1999; 99EP-0123738.  
 xx PR 01-DEC-1998; 98DE-1055312.

XX (DEGS) DEGUSSA-HUELS AG.  
 PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.  
 XX  
 PI Eggeling L, Thierbach G, Sahn H;  
 XX  
 DR WPI: 2000-378263/33.  
 DR P-PSDB: AAB10035.  
 XX  
 PT Recombinant Corynebacterium DNA useful for production of pantothenic  
 PT acid vitamin, comprises panB, panC or livd genes encoding enzymes -  
 XX  
 PS Claim 1c; Page 18-20; 27pp; German.  
 XX  
 CC This invention describes novel recombinant Corynebacterium DNA (I),  
 CC present in microorganisms of the Corynebacterium genus and comprising  
 CC at least one of the panB (ketopantoxydehydrogenase), panC  
 CC (pantothenic acid synthetase), especially the panBC operon, and/or livd  
 CC (dihydroxyaciddehydratase) genes. (I) is useful for the preparation of  
 CC pantothenic acid a vitamin which has applications including cosmetics,  
 CC medicine and human and animal nutrition. The new preparation method using  
 CC fermentation techniques produces the required stereo-isoform D form of  
 CC pantothenic acid. This sequence encodes the Corynebacterium glutamicum  
 CC livd protein which is described in the method of the invention.  
 XX  
 SQ Sequence 2952 BP; 710 A; 933 C; 743 G; 566 T; 0 other;  
 Query Match 82.0%; Score 16.4; DB 21; Length 2952;  
 Best Local Similarity 94.4%; Pred. No. 1.5e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GAACGCGCTCAGAGAAGCC 18  
 |||||  
 DB 2008 GAACGCGCTCAGAGAAGCC 2025  
 RESULT 9  
 AAH68528/c  
 ID AAH68528 standard; DNA; 349980 BP.  
 XX  
 AC AAH68528;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum coding sequence fragment SEQ ID NO: 7063.  
 XX  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis; ds.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW) KYOWA HAKKO KOGYO KK.  
 XX  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI: 2001-376931/40.  
 XX  
 PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT

PS Disclosure; SEQ ID NO: 7063; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX

SQ Sequence 349980 BP; 80727 A; 91049 C; 97618 G; 80586 T; 0 other;  
 Query Match 82.0%; Score 16.4; DB 22; Length 349980;  
 Best Local Similarity 94.4%; Pred. NO. 2.4e+02;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCGCTCAGAGAAGCC 18  
 |||||

DB 133559 GAACGCGCTCAGAGAAGCC 133542

RESULT 10  
 AAH99656/c  
 ID AAH99656 standard; cDNA; 2480 BP.

XX AAH99656;

XX 16-OCT-2001 (first entry)

DE Human protein encoding cDNA sequence SEQ ID NO:491.

XX Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;  
 KW antitaggregant; haemostatic; vulnery; antidiabetic; cytostatic;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autolmmunity;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder; ss.

XX Homo sapiens.

OS WO200153455-A2.

XX 26-JUL-2001.

PF 22-DEC-2000; 2000WO-US35017.

XX 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-457603/49.

XX P-PSDB: AAM25715.

PT Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 XX Claim 1; Page 559-560; 1217pp; English.  
 XX  
 CC AAH99166 to AAH99904 encode the human proteins given in AAH25225 to  
 CC AAH25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
 CC central nervous system; viricide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;  
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,  
 CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.  
 XX  
 SQ Sequence 2480 BP; 373 A; 777 C; 864 G; 466 T; 0 other;  
 Query Match 80.0%; Score 16; DB 22; Length 2480;  
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 GCCTCAGAGAGCCTG 20  
 DB 2255 GCCTCAGAGAGCCTG 2240  
 |||||  
 RESULT 11  
 AAC20810/c  
 ID AAC20810 standard; cDNA; 155 BP.  
 XX  
 AC AAC20810;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 24885.  
 XX  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 XX  
 PR 26-FEB-1999; 9905-0122487.  
 XX  
 XX (GIST ) GENSET.  
 PA  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI; 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 24885; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 SQ Sequence 155 BP; 52 A; 14 C; 50 G; 39 T; 0 other;  
 Query Match 79.0%; Score 15.8; DB 21; Length 155;  
 Best Local Similarity 89.3%; Pred. No. 2.3e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 AACGCCTCAGAGAGCCTG 20  
 DB 92 AACTCTCAGAGAGCCTG 74  
 |||||  
 RESULT 12  
 AAS98541/c  
 ID AAS98541 standard; cDNA; 267 BP.  
 XX  
 AC AAS98541;  
 XX  
 DT 12-MAR-2002 (first entry)  
 XX  
 DE Human protective DNA sequence CNI-00748 open reading frame DNA #21.  
 XX  
 KW Human; protective sequence; cell death; central nervous system; stroke;  
 KW ischaemia; open reading frame; ORF; cerebral herniation; septic embolism;  
 KW cerebral oedema; meningitis; protozoal infection; malaria; CNI-00733; ss;  
 KW metazoal infection; vascular disease; eye; macular degeneration; trauma;  
 KW diabetic retinopathy; epidural haematoma; tumour; degenerative disease;  
 KW nutritional condition; environmental condition; metabolic condition;  
 KW CNI-00736; CNI-00738; CNI-00742; CNI-00748; cancer; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200181361-A1.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 09-APR-2001; 2001WO-US11501.  
 XX  
 PR 11-APR-2000; 2000US-0547938.  
 XX  
 PA (COGE-) COGENT NEUROSCIENCE INC.  
 XX  
 PI Portbury SD, Puranam K, Katz LC, Lo DC, Barney S, Thomas MB;  
 XX  
 DR WPI; 2002-056433/09.  
 DR P-PSDB; AAU73375.  
 XX  
 PT Polypeptides and polynucleotides comprising protective sequences useful  
 PT for preventing, delaying or rescuing a cell from death in disease,  
 PT condition or disorders such as Alzheimer's disease, stroke, tumours,  
 PT trauma  
 XX  
 PS Claim 2; Fig 8U; 228pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide encoded by a protective  
 CC sequence, which is a polynucleotide comprising sequences which, when  
 CC introduced into a cell either predisposed to undergo cell death or in the  
 CC process of undergoing cell death, prevent, delay, or rescue the cell from  
 CC death, relative to a corresponding cell into which exogenous nucleic  
 CC acids have been introduced. The sequences of the invention are useful for

CC diagnosing a protective sequence-mediated condition, disorder or disease  
CC in an individual. The treatable disorders are preferably of the central  
CC nervous system of humans including ischaemia-related conditions such as  
CC stroke, cerebral herniation, septic embolism, cerebral oedema, infections  
CC such as meningitis, protozoal infections such as malaria, metazoal  
CC infections such as echinococcosis, vascular diseases such as ischaemic  
CC encephalopathy, conditions involving the eye such as macular  
CC tumours such as primary intracranial tumours, trauma such as epidural haematoma,  
CC as Alzheimer's disease and nutritional, degenerative diseases such  
CC conditions. Sequences AAS98409-AAS98544 represent human protective  
CC sequence DNA and open reading frames of the polynucleotides.

XX  
SQ Sequence 267 BP; 49 A; 80 C; 85 G; 53 T; 0 other;

Query Match 79.0%; Score 15.8; DB 24; Length 267;  
Best Local Similarity 89.5%; Pred. No. 2.4e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GAAGCCTCAGAGAGCCTG 19  
||||||| ||||| |||||  
DB 102 GAAGCCTGAGAGCCTG 84

RESULT 13  
AAS98409/C  
ID AAF11088 standard; cDNA; 274 BP.

XX AAF11088;

DT 13-MAR-2001 (first entry)

DE Fusarium-venenatum EST SEQ ID NO:3611.

KW Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Fusarium venenatum.

XX WO2000056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

XX (NOVO ) NOVO NORDISK BIOTECH INC.

XX (NOVO ) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX WPI; 2000-594572/56.

XX Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags

PS Claim 86; Page 1643; 3161pp; English.

CC The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be

CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway  
CC engineering. Using ESTs provides several advantages over genomic or  
CC random cDNA clones including elimination of redundancy as one spot on an  
CC array equals one gene or open reading frame, and organisation of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
CC all specifically claimed in the present invention.

XX SQ Sequence 274 BP; 67 A; 64 C; 61 G; 67 T; 15 other;

Query Match 79.0%; Score 15.8; DB 21; Length 274;  
Best Local Similarity 85.0%; Pred. No. 2.4e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GAAGCCTCAGAGAGCCTG 20  
||||| ||||| |||||  
DB 144 GAAGCCTCAGAGCCTG 125

RESULT 14  
AAS98540/C

ID AAS98540 standard; cDNA; 321 BP.

XX AAS98540;

DT 12-MAR-2002 (first entry)

DE Human protective DNA sequence CNI-00748 open reading frame DNA #20.

XX Human; protective sequence; cell death; central nervous system; stroke;  
KW ischaemia; open reading frame; ORF; cerebral herniation; septic embolism;  
KW cerebral oedema; meningitis; protozoal infection; malaria; CNI-00733; ss;  
KW metazoal infection; vascular disease; eye; macular degeneration; trauma;  
KW diabetic retinopathy; epidural haematoma; tumour; degenerative disease;  
KW nutritional condition; environmental condition; metabolic condition;  
KW CNI-00736; CNI-00738; CNI-00742; CNI-00748; cancer; gene therapy.

OS Homo sapiens.

XX WO200181361-A1.

XX 01-NOV-2001.

XX 09-APR-2001; 2001WO-US11501.

XX 11-APR-2000; 2000US-0547938.

XX (COGE-) COGENT NEUROSCIENCE INC.

PI Portbury SD, Puranam K, Katz LC, Lo DC, Barney S, Thomas MB;

XX WPI; 2002-066433/09.

XX P-PSDB; AAU73374.

XX Polypeptides and polynucleotides comprising protective sequences useful  
PT for preventing, delaying or rescuing a cell from death in disease,  
PT condition or disorders such as Alzheimer's disease, stroke, tumours,  
PT trauma

PS Claim 2; Fig 8T; 228pp; English.

XX The invention relates to an isolated polypeptide encoded by a protective  
CC sequence, which is a polynucleotide comprising sequences which, when  
CC introduced into a cell either predisposed to undergo cell death or in the  
CC process of undergoing cell death, prevent, delay, or rescue the cell from  
CC death, relative to a corresponding cell into which exogenous nucleic

CC acids have been introduced. The sequences of the invention are useful for  
 CC diagnosing a protective sequence-mediated condition, disorder or disease  
 CC in an individual. The treatable disorders are preferably of the central  
 CC nervous system of humans including ischaemia-related conditions such as  
 CC stroke, cerebral herniation, septic embolism, cerebral oedema, infections  
 CC such as meningitis, protozoal infections such as malaria, metazoal  
 CC infections such as echinococcosis, vascular diseases such as ischaemic  
 CC encephalopathy, conditions involving the eye such as macular  
 CC degeneration, diabetic retinopathy, trauma such as epidural haematoma,  
 CC tumours such as primary intracranial tumours, degenerative diseases such  
 CC as Alzheimer's disease and nutritional, environmental and metabolic  
 CC conditions. Sequences AAS98409-AAS98544 represent human protective  
 CC sequence DNA and open reading frames of the polynucleotides.

XX Sequence 321 BP; 65 A; 92 C; 101 G; 63 T; 0 other;

Query Match 79.0%; Score 15.8; DB 24; Length 321;  
 Best Local Similarity 89.5%; Pred. No. 2.4e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACGCTCAGAGAGCCT 19  
 ||||| ||||| ||||| |||||  
 DB 156 GAACGCTCAGAGAGCCT 138

## RESULT 15

AAT19761  
 ID AAT19761 standard; cDNA to mRNA; 373 BP.

AC AAT19761;

XX 12-JUL-1996 (first entry)

XX Human gene signature HUMGS00836.

XX Gene signature: messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.

XX Homo sapiens.

XX WO9514772-A1.

XX 01-JUN-1995.

XX 11-NOV-1994; 94WO-JP01916.

XX 12-NOV-1993; 93JP-0355504.

XX (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

XX Matsubara K, Okubo K;

XX WPI; 1995-206931/27.

XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues

PS Claim 1; Page 466; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AAT19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.

XX Sequence 373 BP; 89 A; 71 C; 98 G; 102 T; 13 other;

Query Match 79.0%; Score 15.8; DB 16; Length 373;  
 Best Local Similarity 89.5%; Pred. No. 2.5e+02;  
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGCCTCAGAGAGCCTG 20  
 ||||| ||||| ||||| |||||  
 DB 132 AAGCCTCAGAGAGCCTG 150

Search completed: July 8, 2003, 02:19:04  
 Job time : 127.659 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: July 8, 2003, 01:24:03 ; Search time 27.6098 seconds  
(without alignments)  
222.151 Million cell updates/sec

Title: US-09-647-780A-15  
Perfect score: 20  
Sequence: 1 gaacgctcagagaagcctg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Issued\_Patents\_NA:\*

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.4	82.0	2952	4	US-09-318-794A-4
2	15.8	79.0	652	4	US-09-328-111-505
3	15.4	77.0	4503	2	US-08-770-301A-2
4	15.4	77.0	4503	3	US-09-175-581-2
5	15.2	76.0	477	4	US-08-706-391B-6
6	15.2	76.0	1425	2	US-08-883-515-1
7	15.2	76.0	3980	4	US-09-197-814-1
8	15.2	76.0	3980	4	US-09-197-814-2
9	14.8	74.0	467	4	US-08-927-219-133
10	14.8	74.0	777	4	US-09-161-241-76
11	14.8	74.0	1331	4	US-09-179-221B-3
12	14.8	74.0	3191	4	US-09-453-702B-3
13	14.4	72.0	707	2	US-08-850-910A-70
14	14.4	72.0	1251	1	US-08-265-086-3
15	14.4	72.0	1504	2	US-08-850-910A-17
16	14.4	72.0	1507	2	US-08-850-910A-38
17	14.4	72.0	2896	1	US-08-441-430-31
18	14.4	72.0	2995	1	US-08-441-430-32
19	14.4	72.0	9421	2	US-08-370-319C-2
20	14.4	72.0	9421	2	US-09-224-834-2
21	14.4	72.0	14042	4	US-08-652-877-85
22	14.4	72.0	14044	4	US-08-652-877-89
23	14.4	72.0	14080	4	US-08-652-877-87
24	14.4	72.0	14083	4	US-08-476-515A-83
25	14.4	72.0	14086	4	US-08-652-877-83
26	14.4	72.0	169998	4	US-09-676-610B-24
27	14.2	71.0	960	1	US-08-624-125-13

```

c 28 14.2 71.0 1212 3 US-09-046-578-3 Sequence 3, Appli
29 14.2 71.0 1257 1 US-08-487-748A-11 Sequence 11, Appli
30 14.2 71.0 1257 3 US-08-480-070C-11 Sequence 11, Appli
31 14.2 71.0 1257 3 US-08-829-525-11 Sequence 11, Appli
32 14.2 71.0 1257 4 US-08-609-583A-11 Sequence 11, Appli
33 14.2 71.0 1257 4 US-08-937-393-11 Sequence 11, Appli
34 14.2 71.0 1257 4 US-09-310-367-11 Sequence 11, Appli
35 14.2 71.0 1257 4 US-09-032-337-11 Sequence 11, Appli
36 14.2 71.0 1994 4 US-09-398-395A-41 Sequence 41, Appli
37 14.2 71.0 2275 4 US-09-391-104-1 Sequence 1, Appli
38 14.2 71.0 3509 2 US-08-817-436A-1 Sequence 1, Appli
39 14.2 71.0 3849 4 US-09-202-893B-1 Sequence 1, Appli
40 14.2 71.0 7430 4 US-08-976-259-64 Sequence 64, Appli
41 14.2 71.0 10079 2 US-08-476-866-20 Sequence 20, Appli
42 14.2 71.0 25002 4 US-08-961-527-48 Sequence 48, Appli
43 14.2 71.0 50341 1 US-08-247-901C-1 Sequence 1, Appli
44 14.2 71.0 50341 2 US-09-075-904-1 Sequence 1, Appli
45 14.2 71.0 52297 4 US-09-426-436-1 Sequence 1, Appli

```

#### ALIGNMENTS

```

RESULT 1
; Sequence 4, Application US/09318794A
; Patent No. 6177264
; GENERAL INFORMATION:
; APPLICANT: DEGUSSA AKTIENGESellschaft
; TITLE OF INVENTION: METHOD FOR THE FERMENTATIVE PRODUCTION OF D-PANTOTHENIC
; TITLE OF INVENTION: ACID USING CORYNEFORM BACTERIA
; FILE REFERENCE: eggeling
; CURRENT APPLICATION NUMBER: US/09/318,794A
; PRIOR FILING DATE: 1999-05-26
; PRIOR APPLICATION NUMBER: DE 198 55 312.9
; PRIOR FILING DATE: 1998-12-01
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2952
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (290)..(2125)
; US-09-318-794A-4

```

```

Query Match 82.0%; Score 16.4; DB 4; Length 2952;
Best Local Similarity 94.4%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GAACGCTCAGAGAAGCC 18
      |||||
Db 2008 GAACGCTCAGAGAAGCC 2025

```

```

RESULT 2
; Sequence 505, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert

```

## TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

FILE REFERENCE: CCD-257 (US)  
CURRENT APPLICATION NUMBER: US/09/328,111  
EARLIER FILING DATE: 1999-06-08  
EARLIER APPLICATION NUMBER: US 60/088,801  
EARLIER FILING DATE: 1998-06-10  
NUMBER OF SEQ ID NOS: 850  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 505  
LENGTH: 652

TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(652)  
OTHER INFORMATION: n = A,T,C or G

US-09-328-111-505  
Query Match 79.0%; Score 15.8; DB 4; Length 652;  
Best Local Similarity 89.5%; Pred. No. 32;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACGCTCAGAGAGCCTG 20

Db 314 AACGCCACAGAGAGCCTG 296

## RESULT 3

US-08-770-301A-2  
Sequence 2, Application US/08770301A  
Patent No. 5948637

GENERAL INFORMATION:  
APPLICANT: IKEDA, JUN  
APPLICANT: KANEDA, SUMIKO  
APPLICANT: YANAGI, HIDEKI  
APPLICANT: MATSUMOTO, MASAYASU  
APPLICANT: YURA, TAKASHI  
TITLE OF INVENTION: NOVEL STRESS PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:

ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/770,301A  
FILING DATE: 20-DEC-1996  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR, GERALD M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1422-287  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4503 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
IDENTIFICATION METHOD: E

## US-08-770-301A-2

Query Match 77.0%; Score 15.4; DB 2; Length 4503;  
Best Local Similarity 94.1%; Pred. No. 65;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 CGCCTCAGAGAGCCTG 20

Db 2739 CGCCACAGAGAGCCTG 2755

## RESULT 4

US-09-175-581-2  
Sequence 2, Application US/09175581  
Patent No. 6034232

GENERAL INFORMATION:  
APPLICANT: IKEDA, JUN  
APPLICANT: KANEDA, SUMIKO  
APPLICANT: YANAGI, HIDEKI  
APPLICANT: MATSUMOTO, MASAYASU  
APPLICANT: YURA, TAKASHI  
TITLE OF INVENTION: NOVEL STRESS PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/175,581  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA: US 08/770,301  
APPLICATION NUMBER: US 08/770,301  
FILING DATE: 20-DEC-1996  
ATTORNEY/AGENT INFORMATION:

NAME: MURPHY JR, GERALD M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1422-287  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4503 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:

NAME/KEY: CDS  
IDENTIFICATION METHOD: E  
US-09-175-581-2

Query Match 77.0%; Score 15.4; DB 3; Length 4503;  
Best Local Similarity 94.1%; Pred. No. 65;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 CGCCTCAGAGAGCCTG 20

Db 2739 CGCCACAGAGAGCCTG 2755

## RESULT 5

US-08-706-391B-6/c  
Sequence 6, Application US/08706391B

Patent No. 6174725  
GENERAL INFORMATION:  
APPLICANT: ANDERSON, OLIN D  
TITLE OF INVENTION: ALTERING DOUGH VISCOELASTICITY WITH MODIFIED GLUTENINS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: USDA, AGRICULTURAL RESEARCH SERVICE, PACIFIC WEST  
STREET: 800 BUCHANAN STREET  
CITY: ALBANY  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94710  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/706,391B  
FILING DATE: 30-Aug-1996  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: CONNOR, MARGARET A  
REGISTRATION/DOCKET NUMBER: 30,043  
REFERENCE/DOCKET NUMBER: 0235.95/USDA96-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 559-6067  
TELEFAX: (510) 559-5777  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 477 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-08-706-391B-6  
Query Match 76.0%; Score 15.2; DB 4; Length 477;  
Best Local Similarity 85.0%; Pred. No. 62;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GAACGCTCAGAGAGCCCTG 20  
Db 341 GAACGCTCGGAGAGCCCTG 322  
RESULT 6  
US-08-883-515-1/c  
GENERAL INFORMATION:  
APPLICANT: Austin-Phillips, Sandra  
APPLICANT: Burgess, Richard R.  
APPLICANT: German, Thomas L.  
APPLICANT: Ziegelhoffer, Thomas L.  
TITLE OF INVENTION: Transgenic Plants as an Alternative  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: deWitt Ross & Stevens, S.C.  
STREET: 8000 Excelsior Drive, Ste. 401  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53717-1914  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,495

FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,718  
FILING DATE: 17-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sara, Charles S.  
REGISTRATION NUMBER: 30,492  
REFERENCE/DOCKET NUMBER: 09820,036  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 831-2100  
TELEFAX: (608) 831-2106  
Sequence 1, Application US/08883515  
Patent No. 5981836  
GENERAL INFORMATION:  
APPLICANT: Osteryoung, Katherine W  
TITLE OF INVENTION: PLANT CHLOROPLAST DIVISION GENES  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,515  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Sedy, Nicholas J.  
REGISTRATION/DOCKET NUMBER: 27,386  
REFERENCE/DOCKET NUMBER: 920905,90016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1425 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 25..1326  
US-08-883-515-1  
Query Match 76.0%; Score 15.2; DB 2; Length 1425;  
Best Local Similarity 85.0%; Pred. No. 71;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GAACGCTCAGAGAGCCCTG 20  
Db 1167 GAATGACTGAGAGAGCCCTG 1148  
RESULT 7  
US-09-197-814-1/c  
Sequence 1, Application US/09197814A  
Patent No. 6316220  
GENERAL INFORMATION:  
APPLICANT: Christensen, Tove  
TITLE OF INVENTION: A Transcription Factor  
FILE REFERENCE: 4484,204-US  
CURRENT APPLICATION NUMBER: US/09/197,814A  
CURRENT FILING DATE: 1998-11-23  
EARLIER APPLICATION NUMBER: 0740/96

EARLIER FILING DATE: 1996-07-05  
EARLIER APPLICATION NUMBER: PCT/DK97/00305  
EARLIER FILING DATE: 1997-07-07  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0.  
SEQ ID NO 1  
LENGTH: 3980  
TYPE: DNA  
ORGANISM: Aspergillus oryzae  
US-09-197-814-1

Query Match 76.0%; Score 15.2; DB 4; Length 3980;  
Best Local Similarity 85.0%; Pred. No. 81;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGCCTG 20  
||||| ||| ||||| |||  
DB 2062 GAAGACTCCGAGAGTCTG 2043

RESULT 8  
US-09-197-814-2/c  
Sequence 2, Application US/09197814A  
Patent No. 6316220  
GENERAL INFORMATION:  
APPLICANT: Christensen, Tove  
TITLE OF INVENTION: A Transcription Factor  
FILE REFERENCE: 484.204-US  
CURRENT APPLICATION NUMBER: US/09/197,814A  
CURRENT FILING DATE: 1998-11-23  
EARLIER APPLICATION NUMBER: 0740/96  
EARLIER FILING DATE: 1996-07-05  
EARLIER APPLICATION NUMBER: PCT/DK97/00305  
EARLIER FILING DATE: 1997-07-07  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 3980  
TYPE: DNA  
ORGANISM: Aspergillus oryzae  
US-09-197-814-2

Query Match 76.0%; Score 15.2; DB 4; Length 3980;  
Best Local Similarity 85.0%; Pred. No. 81;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGCCTG 20  
||||| ||| ||||| |||  
DB 2062 GAAGACTCCGAGAGTCTG 2043

RESULT 9  
US-08-927-219-133  
Sequence 133, Application US/08927219  
Patent No. 6187533  
GENERAL INFORMATION:  
APPLICANT: Bell, Graeme I.  
APPLICANT: Yamagata, Kazuya  
APPLICANT: Oda, Naohisa  
APPLICANT: Kaisaki, Pamela J.  
APPLICANT: Furuta, Hiroto  
APPLICANT: Horikawa, Yukio  
APPLICANT: Menzel, Stephen  
TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY  
GENES  
TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA

ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/927,219  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/029,679  
FILING DATE: 30-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/028,056  
FILING DATE: 02-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/025,719  
FILING DATE: 10-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Wilson, Mark B.  
REGISTRATION NUMBER: 37,259  
REFERENCE/DOCKET NUMBER: ARCD:272  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512/418-3000  
TELEFAX: 512/474-7577  
INFORMATION FOR SEQ ID NO: 133:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 467 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-927-219-133

Query Match 74.0%; Score 14.8; DB 4; Length 467;  
Best Local Similarity 88.9%; Pred. No. 98;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGCCTCAGAGCCTG 20  
| | | | | | | | | | | | | | | | | |  
DB 293 ATGCTCAGAGCCTG 310

RESULT 10  
US-09-161-241-76/c  
Sequence 76, Application US/09161241  
Patent No. 6344541  
GENERAL INFORMATION:  
APPLICANT: Bass, Michael B.  
APPLICANT: Sullivan, John K.  
APPLICANT: Theill, Lars E.  
APPLICANT: Wang, Daquang  
TITLE OF INVENTION: NOVEL DKR POLYPEPTIDES  
FILE REFERENCE: A-548  
CURRENT APPLICATION NUMBER: US/09/161,241  
CURRENT FILING DATE: 1998-09-25  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 76  
LENGTH: 777  
TYPE: DNA  
ORGANISM: Human  
US-09-161-241-76

Query Match 74.0%; Score 14.8; DB 4; Length 777;  
Best Local Similarity 88.9%; Pred. No. 1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGCCTG 18  
||||| ||||| ||||| |||||  
DB 20 GAAGCATCAGAGCCTG 3

RESULT 11  
US-09-179-221D-3  
; Sequence 3, Application US/09179221D  
; Patent No. 6291168  
; GENERAL INFORMATION:  
; APPLICANT: Musso, Richard  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES DIAGNOSTIC FOR PATHOGENIC E.  
; FILE REFERENCE: 922.6496P  
; CURRENT APPLICATION NUMBER: US/09/179,221D  
; CURRENT FILING DATE: 1998-10-27  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: No. 6291168epad  
; SEQ ID NO 3  
; LENGTH: 1331  
; TYPE: DNA  
; ORGANISM: E. coli  
US-09-179-221D-3

Query Match 74.0%; Score 14.8; DB 4; Length 1331;  
Best Local Similarity 88.9%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAGCC 18  
||||| ||||| ||||| |||||  
DB 374 GAAGCCTCAAGAGAGCC 391

RESULT 12  
US-09-453-702B-70/c  
; Sequence 70, Application US/09453702B  
; Patent No. 6365723  
; GENERAL INFORMATION:  
; APPLICANT: Blatner, Frederick R.  
; Burland, Valerie  
; Perna, Nicole T.  
; Plunkett, Guy  
; Welch, Rod  
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3191  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 70:  
US-09-453-702B-70

Query Match 74.0%; Score 14.8; DB 4; Length 3191;  
Best Local Similarity 88.9%; Pred. No. 1.2e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAGCC 18  
||||| ||||| ||||| |||||  
DB 275 GAAGCCTCAAGAGAGCC 258

RESULT 13  
US-08-850-910A-40  
; Sequence 40, Application US/08850910A  
; Patent No. 5948761  
; GENERAL INFORMATION:  
; APPLICANT: SEILHAMER, J.J.  
; APPLICANT: LEWICKI, J.  
; APPLICANT: SCARBOROUGH, R.M.  
; TITLE OF INVENTION: RECOMBINANT TECHNIQUES FOR  
; TITLE OF INVENTION: PRODUCTION OF BRAIN NATRIURETIC PEPTIDE  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER, LLP  
; STREET: 2000 Pennsylvania Avenue, NW, Suite 5500  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1888

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/850,910A  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/477,226  
FILING DATE: 08-FEB-1990  
APPLICATION NUMBER: 07/299,880  
FILING DATE: 19-JAN-1989  
APPLICATION NUMBER: 07/206,470  
FILING DATE: 14-JUN-1988  
APPLICATION NUMBER: 07/200,383  
FILING DATE: 31-MAY-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 219002025212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-1500  
TELEFAX: 202-822-0168  
TELEX:

INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 707 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 100...492  
OTHER INFORMATION:  
US-08-850-910A-40

Query Match 72.0%; Score 14.4; DB 2; Length 707;  
Best Local Similarity 93.8%; Pred. No. 1.6e+02;  
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GCCTCAGAGAGCCTG 20

Db 305 GCCTCAGAGGCTG 320

## RESULT 14

US-08-265-086-3

; Sequence 3, Application US/08265086

; Patent No. 5576191

; GENERAL INFORMATION:

; APPLICANT: Gayle, Margit

; APPLICANT: Slack, Jennifer

; APPLICANT: Gruss, Hans-Juergen

; APPLICANT: Sims, John E.

; APPLICANT: Dower, Steven K.

; TITLE OF INVENTION: No. 5576191el Cytokine That Binds ST2

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Macintosh

; OPERATING SYSTEM: Apple 7.1

; SOFTWARE: Microsoft Word, Version 5.1a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/265,086

; FILING DATE: June 17, 1994

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Seese, Kathryn A.

; REGISTRATION NUMBER: 32,172

; REFERENCE/DOCKET NUMBER: 2825

; TELEPHONE: (206) 587-0430

; TELEFAX: (206) 233-0644

; TELEX: 756822

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1251 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; IMMEDIATE SOURCE:

; CLONE: must2-11g

; FEATURE:

; NAME/KEY: mat\_peptide

; LOCATION: 64..672

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..675

; FEATURE:

; NAME/KEY: sig\_peptide

; LOCATION: 1..63

US-08-265-086-3

Query Match

Best Local Similarity 72.0%; Score 14.4; DB 1; Length 1251;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GCCTCAGAGGCTG 20

|||||

Db 1205 GCCTCAGAGGCTG 1220

## RESULT 15

US-08-850-910A-17

; Sequence 17, Application US/08850910A

; Patent No. 5948761

; GENERAL INFORMATION:

; APPLICANT: SEILHAMER, J.J.

; APPLICANT: LEWICKI, J.

; APPLICANT: SCARBOROUGH, R.M.

; TITLE OF INVENTION: RECOMBINANT TECHNIQUES FOR

; PRODUCTION OF BRAIN NATRIURETIC PEPTIDE

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORRISON & FOERSTER, LLP

; STREET: 2000 Pennsylvania Avenue, NW, Suite 5500

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20006-1888

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: Windows

; SOFTWARE: FastSeq for Windows Version 2.0b

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/850,910A

; FILING DATE: 05-MAY-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/477,226

; FILING DATE: 08-FEB-1990

; APPLICATION NUMBER: 07/299,880

; FILING DATE: 19-JAN-1989

; APPLICATION NUMBER: 07/206,470

; FILING DATE: 14-JUN-1988

; APPLICATION NUMBER: 07/200,383

; FILING DATE: 31-MAY-1988

; ATTORNEY/AGENT INFORMATION:

; NAME: Murashige, Kate H

; REGISTRATION NUMBER: 29,959

; REFERENCE/DOCKET NUMBER: 219002025212

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-887-1500

; TELEFAX: 202-822-0168

; TELEX:

; INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1504 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; FEATURE:

; NAME/KEY: Coding Sequence

; LOCATION: 100..630

; OTHER INFORMATION:

US-08-850-910A-17

Query Match

Best Local Similarity 72.0%; Score 14.4; DB 2; Length 1504;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 GCCTCAGAGGCTG 20

|||||

Db 550 GCCTCAGAGGCTG 565

Search completed: July 8, 2003, 09:32:02

Job time : 28.6598 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 115.512 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-15

Perfect score: 20

Sequence: 1 gaacgcctcagagaagcctg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.4	82.0	1839	9	US-09-738-626-1399
c 2	16.4	82.0	3309400	9	US-09-738-626-1399
3	15.8	79.0	234	10	US-09-911-935A-31
4	15.8	79.0	355	9	US-10-071-766-97
c 5	15.8	79.0	396	9	US-09-918-995-7732
6	15.8	79.0	401	9	US-09-946-807-837
7	15.8	79.0	401	10	US-09-795-668-837
8	15.8	79.0	401	10	US-09-795-668-837
9	15.8	79.0	411	9	US-09-918-995-17753
c 10	15.8	79.0	442	10	US-09-864-761-4088
11	15.8	79.0	453	9	US-09-736-457-828
12	15.8	79.0	453	9	US-09-902-941-828
13	15.8	79.0	453	9	US-09-849-626-828
14	15.8	79.0	453	9	US-10-017-754-828
c 15	15.8	79.0	454	9	US-09-918-995-2506
c 16	15.8	79.0	478	9	US-10-042-125A-26
17	15.8	79.0	492	9	US-10-046-935-1702
18	15.8	79.0	492	9	US-09-878-178-1702
19	15.8	79.0	492	9	US-10-146-502-1702

Sequence 20258, A  
Sequence 505, App  
Sequence 626, App  
Sequence 286, App  
Sequence 364, App  
Sequence 179, App  
Sequence 300, App  
Sequence 601, App  
Sequence 1351, App  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 15509, A  
Sequence 93, Appl  
Sequence 734, App  
Sequence 6543, App  
Sequence 804, App  
Sequence 1876, App  
Sequence 224, App  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 1282, App  
Sequence 1922, App  
Sequence 9556, App

20 15.8 79.0 503 9 US-09-918-995-20258  
c 21 15.8 79.0 652 10 US-09-879-536-505  
22 15.8 79.0 1533 10 US-09-887-576-626  
23 15.8 79.0 1958 10 US-09-925-302-286  
24 15.8 79.0 1959 9 US-10-106-698-364  
25 15.8 79.0 2019 9 US-09-919-039-179  
26 15.8 79.0 2062 9 US-10-203-823-300  
c 27 15.8 79.0 3198 9 US-10-152-661-601  
c 28 15.8 79.0 3198 9 US-09-866-050A-601  
29 15.8 79.0 8841 9 US-09-954-531-1351  
30 15.8 79.0 1503841 9 US-09-946-807-1  
31 15.8 79.0 1503841 10 US-09-795-668-1  
32 15.8 79.0 1503841 10 US-09-795-668-1  
33 15.4 77.0 458 10 US-09-864-761-15509  
34 15.4 77.0 4521 12 US-10-002-600-93  
35 15.2 76.0 348 10 US-09-925-381-1  
36 15.2 76.0 1026 9 US-10-156-761-6543  
c 37 15.2 76.0 1254 9 US-10-037-270-804  
c 38 15.2 76.0 1302 9 US-09-938-842A-1876  
c 39 15.2 76.0 1958 10 US-09-974-300-224  
c 40 15.2 76.0 3980 10 US-09-920-581-1  
c 41 15.2 76.0 3980 10 US-09-920-581-1  
c 42 15.2 76.0 13873 9 US-09-764-868-1282  
c 43 15.2 76.0 15734 9 US-10-091-504-1922  
c 44 15.2 76.0 15734 10 US-09-764-869-1922  
c 45 15.2 76.0 31304 9 US-09-764-891-9556

#### ALIGNMENTS

RESULT 1  
US-09-738-626-1399  
; Sequence 1399, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIALI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09738.626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1399  
; LENGTH: 1839  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1399

Query Match 82.0%; Score 16.4; DB 9; Length 1839;  
Best Local Similarity 94.4%; Pred. No. 42;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCTCAGAGAAGCC 18  
Db 1722 GAACGCTCCGAGAAGCC 1739

RESULT 2  
US-09-738-626-1/c  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: MIZOGUCHI, SATOSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 82.0%; Score 16.4; DB 9; Length 3309400;  
Best Local Similarity 94.4%; Pred. No. 29;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAAGCGCTCAGAGAGCC 18  
Db 1333559 GAAGCGCTCAGAGAGCC 1333542

RESULT 3  
US-09-911-935A-31  
; Sequence 31, Application US/09911935A  
; Patent No. US20020081611A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Brien, Thomas  
; APPLICANT: Guo, Yong Jun  
; TITLE OF INVENTION: ODC Allelic Analysis Method For Assessing Carcinogenic Susceptibility  
; FILE REFERENCE: 9855-3202  
; CURRENT APPLICATION NUMBER: US/09/911,935A  
; CURRENT FILING DATE: 2001-07-24  
; PRIOR APPLICATION NUMBER: US 60/122,301  
; PRIOR FILING DATE: 1999-03-01  
; PRIOR APPLICATION NUMBER: US 09/516,357  
; PRIOR FILING DATE: 2000-03-01  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 31  
; LENGTH: 234  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc-feature  
; LOCATION: (86)...(86)  
; OTHER INFORMATION: n - c or t  
US-09-911-935A-31

Query Match 79.0%; Score 15.8; DB 10; Length 234;  
Best Local Similarity 89.5%; Pred. No. 94;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AACGCGCTCAGAGAGCCCTG 20  
Db 160 AACGCCACAGAGAGCCCTG 178

RESULT 4  
US-10-071-766-97  
; Sequence 97, Application US/10071766  
; Publication No. US20020192678A1  
; GENERAL INFORMATION:  
; APPLICANT: Huel-Mei Chen  
; TITLE OF INVENTION: GENES EXPRESSED IN SENESENCE  
; FILE REFERENCE: PA-0043 US  
; CURRENT APPLICATION NUMBER: US/10/071,766  
; CURRENT FILING DATE: 2002-02-07  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: PERL Program  
; SEQ ID NO 97  
; LENGTH: 355  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc-feature  
; OTHER INFORMATION: Incyte ID No. US20020192678A1 1398377.1  
US-10-071-766-97

Query Match 79.0%; Score 15.8; DB 9; Length 355;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AACGCGCTCAGAGAGCCCTG 20  
Db 72 AACGCCACAGAGAGCCCTG 90

RESULT 5  
US-09-918-995-7732/c  
; Sequence 7732, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7732  
; LENGTH: 396  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-7732

Query Match 79.0%; Score 15.8; DB 9; Length 396;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAAGCGCTCAGAGAGCCCT 19  
Db 238 GAAGCGCTCAGAGAGCCCT 220

RESULT 6  
US-09-946-807-837  
; Sequence 837, Application US/09946807  
; Patent No. US20020165144A1  
; GENERAL INFORMATION:  
; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinhorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.

; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001  
; CURRENT APPLICATION NUMBER: US/09/946,807  
; CURRENT FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: US/09/195,668  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 837  
; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-946-807-837

Query Match 79.0%; Score 15.8; DB 9; Length 401;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAAGCCT 19  
||| || ||||| ||||| |||||  
Db 160 GAAAGCTTCAGAGAAGCCT 178

## RESULT 7

US-09-795-668-837  
; Sequence 837, Application US/09795668  
; Patent No. US20020045377A1  
; GENERAL INFORMATION:

; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2004-001

; CURRENT APPLICATION NUMBER: US/09/795,668  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,716  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 837  
; LENGTH: 401  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-795-668-837

Query Match 79.0%; Score 15.8; DB 10; Length 401;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAAGCCT 19  
||| || ||||| ||||| |||||  
Db 160 GAAAGCTTCAGAGAAGCCT 178

## RESULT 8

US-09-795-686-837  
; Sequence 837, Application US/09795686  
; Patent No. US20020094954A1  
; GENERAL INFORMATION:

; APPLICANT: Stefansson, Hreinn  
; APPLICANT: Steinthorsdottir, Valgerdur  
; APPLICANT: Gulcher, Jeffrey R.  
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
; FILE REFERENCE: 2345.2005-001

; CURRENT APPLICATION NUMBER: US/09/795,686  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 09/515,715  
; PRIOR FILING DATE: 2000-02-28  
; NUMBER OF SEQ ID NOS: 1531

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 837

; LENGTH: 401  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-795-686-837

Query Match 79.0%; Score 15.8; DB 10; Length 401;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAAGCCT 19  
||| || ||||| ||||| |||||  
Db 160 GAAAGCTTCAGAGAAGCCT 178

## RESULT 9

US-09-918-995-17753  
; Sequence 17753, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17753  
; LENGTH: 411  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-918-995-17753

Query Match 79.0%; Score 15.8; DB 9; Length 411;  
Best Local Similarity 89.5%; Pred. No. 92;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACGCCCTCAGAGAAGCCTG 20  
||| || ||||| ||||| |||||  
Db 199 AACGCCACAGAGCAGCCTG 217

## RESULT 10

US-09-864-761-4088/c  
; Sequence 4088, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F  
; FILE REFERENCE: Aesmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 4088  
;; LENGTH: 442  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC007249.2  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.3  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7  
;; US-09-864-761-4088

Query Match 79.0%; Score 15.8; DB 10; Length 442;  
Best Local Similarity 89.5%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AACGCCCTCAGAGAGCCTG 20  
||||| ||||| ||||| |||||  
Db 346 AACGCCACAGAGCCTG 328

RESULT 11  
US-09-736-457-828  
;; Sequence 828, Application US/09736457  
;; Patent No. US20020168637A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, Tongtong  
;; APPLICANT: Bangur, Chaitanya S.  
;; APPLICANT: Lodes, Michael A.  
;; APPLICANT: Fanger, Gary  
;; APPLICANT: Vedvick, Tom  
;; APPLICANT: Carter, Darrick  
;; APPLICANT: Retter, Marc  
;; APPLICANT: Mannion, Jane  
;; APPLICANT: Fan, Liqun  
;; APPLICANT: Wang, Aijun  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
;; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.478C15  
;; CURRENT APPLICATION NUMBER: US/09/736,457  
;; CURRENT FILING DATE: 2000-12-13  
;; NUMBER OF SEQ ID NOS: 1864  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 828

;; LENGTH: 453  
;; TYPE: DNA  
;; ORGANISM: Homo sapien  
;; US-09-736-457-828

Query Match 79.0%; Score 15.8; DB 9; Length 453;  
Best Local Similarity 89.5%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AACGCCCTCAGAGAGCCTG 20  
||||| ||||| ||||| |||||  
Db 374 AACGCCACAGAGCCTG 392

RESULT 12  
US-09-902-941-828  
;; Sequence 828, Application US/09902941  
;; Patent No. US20020172952A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Henderson, Robert A.  
;; APPLICANT: Wang, Tongtong  
;; APPLICANT: Watanabe, Yoshihiro  
;; APPLICANT: Johnson, Jeffrey C.  
;; APPLICANT: Retter, Marc W.  
;; APPLICANT: Marnerakis, Margarita  
;; APPLICANT: Carter, Darrick  
;; APPLICANT: Fanger, Gary R.  
;; APPLICANT: Vedvick, Thomas S.  
;; APPLICANT: Bangur, Chaitanya S.  
;; APPLICANT: McNabb, Andrea  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
;; FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.478C17  
;; CURRENT APPLICATION NUMBER: US/09/902,941  
;; CURRENT FILING DATE: 2001-07-10  
;; NUMBER OF SEQ ID NOS: 2002  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 828  
;; LENGTH: 453  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; US-09-902-941-828

Query Match 79.0%; Score 15.8; DB 9; Length 453;  
Best Local Similarity 89.5%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 2 AACGCCCTCAGAGAGCCTG 20  
||||| ||||| ||||| |||||  
Db 374 AACGCCACAGAGCCTG 392

RESULT 13  
US-09-849-626-828  
;; Sequence 828, Application US/09849626  
;; Publication No. US20020197669A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Bangur, Chaitanya  
;; APPLICANT: Fanger, Gary  
;; APPLICANT: Wang, Aijun  
;; APPLICANT: Wang, Tongtong  
;; APPLICANT: Switzer, Anne  
;; APPLICANT: McNeill, Patricia  
;; APPLICANT: Clapper, Jonathan  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
;; FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER  
;; FILE REFERENCE: 210121.478C16  
;; CURRENT APPLICATION NUMBER: US/09/849,626  
;; CURRENT FILING DATE: 2001-05-03  
;; NUMBER OF SEQ ID NOS: 1926  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 828  
;; LENGTH: 453

; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-849-626-828

Query Match. 79.0%; Score 15.8; DB 9; Length 453;  
Best Local Similarity 89.5%; Pred. No. 91;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACGCCCTCAGAGAGCCTG 20  
||||| ||||| |||||  
Db 374 AACGCCACAGAGCAGCCTG 392

## RESULT 14.

US-10-017-754-828  
; Sequence 828, Application US/10017754  
; Publication No. US20030054363A1

## ; GENERAL INFORMATION:

; APPLICANT: Henderson, Robert A.  
; APPLICANT: Wang, Tongtong  
; APPLICANT: Watanabe, Yoshihiro  
; APPLICANT: Johnson, Jeffrey C.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Marnerakis, Margarita  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Bangur, Chaitanya S.  
; APPLICANT: McNabb, Andria

## ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

## ; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER

; FILE REFERENCE: 210121.478C18

; CURRENT APPLICATION NUMBER: US/10/017,754

; CURRENT FILING DATE: 2001-10-29

; NUMBER OF SEQ ID NOS: 2004

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 828

; LENGTH: 453

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-017-754-828

## Query Match

Best Local Similarity 79.0%; Score 15.8; DB 9; Length 453;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AACGCCCTCAGAGAGCCTG 20  
||||| ||||| |||||  
Db 374 AACGCCACAGAGCAGCCTG 392

## RESULT 15

US-09-918-995-2506/c

; Sequence 2506, Application US/09918995

; Publication No. US20030073623A1

## ; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

## ; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

## ; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; CURRENT FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2506

; LENGTH: 454

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(454)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1013.95 seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-15  
Perfect score: 20  
Sequence: 1 gaacgctcagagagcgtg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estopl: \*  
7: em\_estro: \*  
8: em\_htc: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_htc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vrt: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rod: \*

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	507	10	BE106100
2	18.4	92.0	378	9	AA146423
3	18.4	92.0	388	9	AI647353
4	18.4	92.0	704	12	BE153322
5	17.4	87.0	516	17	AQ209617
6	17.4	87.0	574	17	AZ223089

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	17.4	87.0	944	12	BG121418
8	17.4	87.0	2091	12	BG847865
9	16.8	84.0	182	10	AW445429
10	16.8	84.0	227	13	BQ992210
c 11	16.8	84.0	252	14	BQ375929
12	16.8	84.0	293	17	AQ100155
c 13	16.8	84.0	365	12	BG382418
14	16.8	84.0	375	9	AL775847
15	16.8	84.0	384	10	AW486108
16	16.8	84.0	400	17	AQ147395
17	16.8	84.0	438	17	AQ173747
c 18	16.8	84.0	444	12	BG661129
c 19	16.8	84.0	447	12	BG661134
c 20	16.8	84.0	454	12	BG382419
21	16.8	84.0	472	17	AQ169284
c 22	16.8	84.0	481	9	AA226901
23	16.8	84.0	494	10	BE233903
24	16.8	84.0	495	17	AQ183015
c 25	16.8	84.0	500	9	AI150181
c 26	16.8	84.0	502	17	AQ125886
c 27	16.8	84.0	506	17	B87963
28	16.8	84.0	529	9	AL644837
29	16.8	84.0	530	10	BE233908
30	16.8	84.0	534	17	AQ252696
31	16.8	84.0	555	17	AQ423146
32	16.8	84.0	577	13	BI776184
33	16.8	84.0	615	12	BG513312
34	16.8	84.0	651	9	AL639255
35	16.8	84.0	693	9	AL652033
c 36	16.8	84.0	700	17	AQ162875
c 37	16.8	84.0	850	12	BG460753
c 38	16.8	84.0	873	14	BQ438223
39	16.8	84.0	929	12	BG290567
40	16.8	84.0	1796	12	BF683707
c 41	16.4	82.0	350	9	AI372676
42	16.4	82.0	451	17	AQ175266
c 43	16.4	82.0	462	12	BF291638
c 44	16.4	82.0	464	17	AQ313589
c 45	16.4	82.0	582	17	AZ830493

#### ALIGNMENTS

RESULT 1  
BE106100  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT

BE106100 507 bp mRNA linear EST 13-JUN-2000  
UI-R-BOL-ask-h-09-0-UI.s1 UI-R-BOL Rattus norvegicus cDNA clone  
UI-R-BOL-ask-h-09-0-UI 3', mRNA sequence.  
BE106100  
EST  
BE106100.1 GI:8498202  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

1 (bases 1 to 507)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
9704477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.weeg.uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized medulla library cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

#### FEATURES

source

Location/Qualifiers

1..507  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-B01-ask-h-09-0-UI"  
/clone\_lib="UI-R-B01"  
/dev\_stage="adult"  
/lab\_host="DHL08 (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; The library (UI-R-B01) is a subtracted library derived from a mixture of the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at [ratat.eng.uiowa.edu](http://ratat.eng.uiowa.edu). The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
TAG\_LIB="UI-R-B01"  
TAG\_TISSUE="medulla"  
TAG\_SEQ="GAACCG"

BASE COUNT 105 a 133 c 158 g 111 t

#### ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 507;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAACGCTCAGAGAGCCCTG 20  
|||||  
DB 372 GAACGCTCAGAGAGCCCTG 391

#### RESULT 2

AA146423/c  
LOCUS  
DEFINITION  
AA146423  
Accession  
Version  
Keywords  
Source  
Organism  
Reference  
Authors  
Title  
Journal  
Comment

AA146423 378 bp mRNA linear EST 11-FEB-1997  
mr66e08.i1 Stratagene mouse testis (#937308) Mus musculus cDNA  
clone IMAGE:602438 5' similar to WP:F26G1.6 CE02698 NEPRILYSIN ;  
mRNA sequence.

AA146423.1 GI:1715811  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 378)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

The WashU-HMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
MGI:367870

seq primer: -40RP from Gibco.  
Location/Qualifiers  
1..388  
/organism="Mus musculus"  
/strain="CD-1"  
/db\_xref="taxon:10090"

#### FEATURES

source

Location/Qualifiers

1..378  
/organism="Mus musculus"  
/strain="CD-1"  
/db\_xref="taxon:10090"  
/clone="IMAGE:602438"  
/clone\_lib="Stratagene mouse testis (#937308)"  
/sex="males"  
/tissue\_type="testis"  
/dev\_stage="10-12 week old"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: testis; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3"

BASE COUNT 76 a 131 c 106 g 65 t

#### ORIGIN

Query Match 92.0%; Score 18.4; DB 9; Length 378;  
Best Local Similarity 95.0%; Pred. No. 2.4e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GAACGCTCAGAGAGCCCTG 20  
|||||  
DB 31 GAATGCTCAGAGAGCCCTG 12

#### RESULT 3

AI647353/c  
LOCUS  
DEFINITION  
AI647353  
Accession  
Version  
Keywords  
Source  
Organism  
Reference  
Authors  
Title  
Journal  
Comment

AI647353 388 bp mRNA linear EST 15-MAR-2000  
mr66e08.y1 Stratagene mouse testis (#937308) Mus musculus cDNA  
clone IMAGE:602438 5' similar to WP:F26G1.6 CE02698 NEPRILYSIN ;  
mRNA sequence.

AI647353.1 GI:4725828  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 388)  
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [mouseest@watson.wustl.edu](mailto:mouseest@watson.wustl.edu)

This clone is available royalty-free through LNL; contact the IMAGE Consortium ([info@image.llnl.gov](http://info@image.llnl.gov)) for further information.  
This read is a RESEQUENCE of a previously sequenced mouse clone.  
This read has been verified (found to hit its original self in the correct orientation)  
Putative full length read  
vector to vector length is 389  
MGI:367870

seq primer: -40RP from Gibco.  
Location/Qualifiers  
1..388  
/organism="Mus musculus"  
/strain="CD-1"  
/db\_xref="taxon:10090"

```

/clone-"IMAGE:602438"
/clone_lib-"Stratagene mouse testis (#937308)"
/sex-"males"
/tissue_type-"testis"
/dev_stage-"10-12 week old"
/lab_host-"SOLR (kanamycin resistant)"
/ECORI; Site.2: XhoI; Vector: pBluescript SK-; Site.1:
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATCGGCACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'
BASE COUNT      82 a 131 c 108 g 67 t
ORIGIN
Query Match      92.0%; Score 18.4; DB 9; Length 388;
Best Local Similarity 95.0%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAACGCTCAGAGAGCCTG 20
Db 31 GAATGCTCAGAGAGCCTG 12

RESULT 4
LOCUS      BG153322      704 bp mRNA linear EST 05-FEB-2001
DEFINITION nag12d11.x1 NCI CGAP_Emb2 Rattus norvegicus cDNA clone
IMAGE:4176045 3' similar to TR:Q9Q2V6 Q9Q2V6 SOLUBLE SECRETED
ENDOPEPTIDASE DELTA. [1] ;contains PTR5.b2 PTR5 repetitive element
// mRNA sequence.
ACCESSION   BG153322
VERSION     BG153322.1 GI:12665352
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 704)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
JOURNAL     Contact: Robert Strausberg, Ph.D.
COMMENT     Email: cgabsr@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library
            Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 422.
FEATURES
            Location/Qualifiers
            1..704
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"
            /clone="IMAGE:4176045"
            /clone_lib="NCI CGAP_Emb2"
            /dev_stage="embryo"
            /lab_host="PH10B (T1 phage-resistant)"
            /note="Organ: embryo, pluripotent cell line; Vector:
            pCMV-SPORT6; Site.1: NotI; Site.2: SalI; Cloned
            unidirectionally. Primer: Oligo dt. Average insert size
            1.54 kb. Constructed by Life Technologies."
BASE COUNT      135 a 197 c 217 g 155 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 12; Length 704;
Best Local Similarity 95.0%; Pred. No. 3.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 GAACGCTCAGAGAGCCTG 20
Db 355 GAATGCTCAGAGAGCCTG 374

RESULT 5
LOCUS      AO209617      516 bp DNA linear GSS 18-SEP-1998
DEFINITION HS-3236_A2_E07_T7 CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone :Plate=3236 Col=14 Row=I, DNA sequence.
ACCESSION   AO209617
VERSION     AO209617.1 GI:3622352
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 516)
            Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
            Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
JOURNAL     99380589
MEDLINE
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3236 row: I column: 14
            Class: BAC ends
            High quality sequence stop: 516.
FEATURES
            Location/Qualifiers
            1..516
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="Plate=3236 Col=14 Row=I"
            /clone_lib="CIT Approved Human Genomic Sperm Library D"
            /sex="male"
            /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
            E-Coli DH10B"
BASE COUNT      155 a 117 c 95 g 140 t 9 others
ORIGIN

Query Match      87.08; Score 17.4; DB 17; Length 516;
Best Local Similarity 90.0%; Pred. No. 8.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACGCTCAGAGAGCCTG 20
Db 415 GAACGCTCAGAGAGTCTG 434

RESULT 6
LOCUS      AZ223089      574 bp DNA linear GSS 14-JUN-2000
DEFINITION RPCI-23-94H20.TJ RPCI-23 Mus musculus genomic clone RPCI-23-94H20,
            DNA sequence.
ACCESSION   AZ223089
VERSION     AZ223089.1 GI:8531138
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 574)
            Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.

```

## TITLE Mouse BAC End Sequences from Library RPCI-23

JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: RPCI-23-94H20-TV  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhaot@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Peter de Jong (plet@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics ([inforesgen.com](http://inforesgen.com)). BAC end page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 94 row: H. column: 20  
 Seq primer: SP6  
 Class: BAC ends

## FEATURES source

Location/Qualifiers  
 1. .574  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-94H20"  
 /clone\_lib="RPCI-23"  
 /sex="Female"  
 /lab\_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methyase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."  
 157 a 130 c 149 g 138 t

## BASE COUNT

ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 574;

Best Local Similarity 94.7%; Pred. No. 8.8e-02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACGCTCAGAGAGCCCTG 20

Db 316 AACTCCTCAGAGAGCCCTG 334

## RESULT 7

BG121418  
 LOCUS 602352942F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4450906 5',  
 DEFINITION mRNA sequence.

ACCESSION BG121418

VERSION BG121418.1 GI:12614927

KEYWORDS EST

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo..

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM10237 row: d column: 11

## FEATURES source

High quality sequence stop: 647.

Location/Qualifiers

1. .944

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4450906"

/clone\_lib="NIH\_MGC\_90"

/tissue\_type="adenocarcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pCMV-Sport6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.

Average insert size 1.7 kb. Library enriched for

full-length clones and constructed by Life Technologies.

Note: this is a NIH\_MGC Library."

242 a 201 c 259 g 242 t

## BASE COUNT

ORIGIN

Query Match 87.0%; Score 17.4; DB 12; Length 944;

Best Local Similarity 94.7%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACGCTCAGAGAGCCCTG 20

Db 648 AACGCCACAGAGAGCCCTG 666

## RESULT 8

BG847865

LOCUS

DEFINITION

1024019E02.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II

Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BG847865

VERSION BG847865.1 GI:14229049

KEYWORDS EST

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.

AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Iefebvre,P.,

McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants: project phase 2

Unpublished (2000)

COMMENT Contact: Charles Hauser

DCMB Box 91000

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: [chauser@duke.edu](mailto:chauser@duke.edu).

Location/Qualifiers

1. .2091

/organism="Chlamydomonas reinhardtii"

/strain="CC-1690 wild type mt+ 21gr"

/db\_xref="taxon:3055"

/clone\_lib="C. reinhardtii CC-1690, normalized, Lambda Zap

II"

/note="Vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:

XhoI; This library, constructed by John Davies and Jeffrey

McDermott, combines cDNAs from CC-1690 cells grown to

mid-log phase in TAP (acetate-containing) medium in the

light, TAP medium in the dark, HS (minimal) medium in

ambient levels of CO2 and HS medium bubbled with 5% CO2.

polyA mRNA was purified from each sample, pooled and cDNA

synthesized. The cDNA was directionally cloned into lambda

ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.

pluescript II SK- plasmids were excised from the lambda

ZAP clones by superinfection with EXassist (Stratagene)

phage. The library was normalized using method 4 described

in Ronaldo et al (1996) Genome Research 6: 791-806."

535 a 494 c 700 g 221 t 141 others

BASE COUNT



Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
MEDLINE  
COMMENT

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-TN0008-010800-123-g09&t3=2000-08-01&t4=1)

Seq primer: puc 18 forward

High quality sequence stop: 21.

Location/Qualifiers

1. .252

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="TN0008"

/dev\_stage="Adult"

/note="Organ: testis, normal; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions.

63 a 42 c 65 g 82 t

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 14; Length 252;

Best Local Similarity 90.0%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAGCCTG 20

||||| ||||| ||||| |||||

Db 25 GAAGCCTCAGAGTAGTCTG 6

||||| ||||| ||||| |||||

RESULT 12

AQ100155

LOCUS

DEFINITION

HS.3049\_B2.D07\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3049 Col=14 Row=H, DNA sequence.

ACCESSION

AQ100155

VERSION

AQ100155.1 GI:3471184

SOURCE

GSS.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 293)

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Query Match 84.0%; Score 16.8; DB 12; Length 365;

Best Local Similarity 90.0%; Pred. No. 1.3e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAGCCTG 20

||||| ||||| ||||| |||||

Db 25 GAAGCCTCAGAGTAGTCTG 6

||||| ||||| ||||| |||||

RESULT 13

BG382418/C

LOCUS

DEFINITION

298310 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.

ACCESSION

BG382418

VERSION

BG382418.1 GI:13306890

KEYWORDS

EST.

SOURCE

pig.

ORGANISM

Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

1 (bases 1 to 365)

Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred v0.980504 e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGAG

Plate: 4 row: D column: 1.

Seq primer: ATTTAGTGCACACTATAG.

Location/Qualifiers

1. .365

/organism="Sus scrofa"

/db\_xref="taxon:9823"

/clone\_lib="MARC lPIG"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;

Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."

74 a 109 c 130 g 52 t

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 12; Length 365;

Best Local Similarity 90.0%; Pred. No. 1.3e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAGCCTG 20

||||| ||||| ||||| |||||

Db 263 GAAGCCTCAAGAGAGACTG 282

||||| ||||| ||||| |||||

FEATURES

source

1. .293

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="Plate=3049 Col=14 Row=H"

/clone\_lib="CIT Approved Human Genomic Sperm Library D"

/sex="male"

/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

92 a 65 c 54 g 81 t 1 others

BASE COUNT

ORIGIN

Query Match 84.0%; Score 16.8; DB 17; Length 293;

Best Local Similarity 90.0%; Pred. No. 1.2e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAAGCCTCAGAGAGCCTG 20

||||| ||||| ||||| |||||

Db 263 GAAGCCTCAAGAGAGACTG 282

||||| ||||| ||||| |||||

QY 1 GAACGCCTCAGAGAGCCTG 20  
 I | | | | | | | | | | | | | | | | |  
 Db 213 GCACGCCTCAGAGAGCCAG 194

## RESULT 14

AL775847

LOCUS

DEFINITION

AL775847

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TNeu073n20.plcSP6

Sequencing primer: PICSP6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

Location/Qualifiers

1. .375

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone\_lib="TNeu073n20"

/clone\_lib="XGC-neurula"

/dev\_stage="neurula"

/lab\_host="Escherichia coli DH10B"

/note="Vector: pCSI07; Site\_1: EcoRI; Site\_2: NotI; CDNA

was oligo dt primed from 5'ug of poly A+ RNA from neurula.

EcoRI-NotI cut cDNA was then ligated into pCSI07 with

EcoRI at the 5' end and NotI at the 3' end."

70 a 108 c 116 g 81 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

18; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

QY 1 GAACGCCTCAGAGAGCCTG 20

I | | | | | | | | | | | | | | | | |

Db 169 GAACGCCTCAGAGAGCCAG 188

RESULT 15

AW486108

LOCUS

DEFINITION

70286 MARC 4BOV Bos taurus

CDNA 5', mRNA sequence.

AW486108

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Smith, T.P.L., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett

G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,

Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and

AL775847 375 bp mRNA linear EST 25-JUN-2002  
 AL775847 XGC-neurula Silurana tropicalis cDNA clone TNeu073n20 5',  
 mRNA sequence.  
 AL775847  
 AL775847  
 AL775847 1 GI:21561551  
 EST.  
 SOURCE western clawed frog.  
 ORGANISM Silurana tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 Xenopodinae; Silurana.  
 1 (bases 1 to 375)

Taylor R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.  
 Sanger Xenopus tropicalis EST project 2002  
 Unpublished (2001)  
 Contact: Taylor R

Sanger Centre  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS\_SEQUENCE\_ID: TNeu073n20.plcSP6  
 Sequencing primer: PICSP6

This sequence is from a Xenopus Gene Collection (XGC) library  
 constructed by Aaron M. Zorn.

Location/Qualifiers  
 1. .375

/organism="Silurana tropicalis"  
 /db\_xref="taxon:8364"  
 /clone\_lib="TNeu073n20"

/clone\_lib="XGC-neurula"  
 /dev\_stage="neurula"  
 /lab\_host="Escherichia coli DH10B"

/note="Vector: pCSI07; Site\_1: EcoRI; Site\_2: NotI; CDNA  
 was oligo dt primed from 5'ug of poly A+ RNA from neurula.  
 EcoRI-NotI cut cDNA was then ligated into pCSI07 with

EcoRI at the 5' end and NotI at the 3' end."  
 70 a 108 c 116 g 81 t

Query Match 84.0%; Score 16.8; DB 9; Length 375;  
 Best Local Similarity 90.0%; Pred. No. 1.4e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACGCCTCAGAGAGCCTG 20  
 I | | | | | | | | | | | | | | | | |

Db 169 GAACGCCTCAGAGAGCCAG 188

RESULT 15  
 AW486108 384 bp mRNA linear EST 25-APR-2001  
 70286 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.

AW486108  
 AW486108 1 GI:7056214  
 EST.

SOURCE cow.  
 ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.

1 (bases 1 to 384)  
 Smith, T.P.L., Freking, B.A., Roberts, A.J., Stone, R.T.,

Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett

G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,

Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and

Keele, J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA  
 libraries and construction of a gene index for cattle  
 genome res. 11 (4), 626-630 (2001)  
 21180013  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 20  
 and -mismatch 12 options.

PCR Primers  
 FORWARD: AGGAACAGCAGTATGACCAT  
 BACKWARD: GTTTCCTCCAGTCACGACG  
 Plate: 37 row: H column: 7  
 Seq primer: ATTAGTGACACTATAG.

FEATURES  
 source

1. .384  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 4BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"

/note="Vector: pCMV SPORT6; Site\_1: NotI; Site\_2: SalI;  
 Library made from pooled tissue from day 20 and day 40  
 embryos."  
 63 a 134 c 108 g 79 t

BASE COUNT  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 384;  
 Best Local Similarity 90.0%; Pred. No. 1.4e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GAACGCCTCAGAGAGCCTG 20  
 I | | | | | | | | | | | | | | | | |

Db 321 GAACGCCTCTGGGAGCCTG 340

Search completed: July 8, 2003, 09:21:56  
 Job time: 1019.1 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:45:53 ; Search time 220.098 seconds  
(without alignments)  
2644.537 Million cell updates/sec

Title: US-09-647-780A-16  
Perfect score: 20  
Sequence: 1 atgaccagaactccagccgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	20	100.0	20	6	AX014716	AX014716 Sequence
2	20	100.0	2765	6	AX014701	AX014701 Sequence
3	20	100.0	174953	2	AC094732	AC094732 Rattus no
4	18.4	92.0	2076	6	AX146976	AX146976 Sequence
5	18.4	92.0	2232	6	AX319864	AX319864 Sequence
6	18.4	92.0	2262	6	AX146980	AX146980 Sequence
7	18.4	92.0	2340	6	AX146978	AX146978 Sequence
8	18.4	92.0	2340	6	AX473102	AX473102 Sequence
9	18.4	92.0	2583	10	AF302075	AF302075 Mus muscu
10	18.4	92.0	2601	10	AF157106	AF157106 Mus muscu
11	18.4	92.0	2636	6	AX139743	AX139743 Sequence
12	18.4	92.0	2652	10	AF302076	AF302076 Mus muscu
13	18.4	92.0	2663	6	AX139745	AX139745 Sequence
14	18.4	92.0	2676	6	AX033274	AX033274 Sequence
15	18.4	92.0	2694	10	AF302077	AF302077 Mus muscu
16	18.4	92.0	2714	6	AX139741	AX139741 Sequence
17	18.4	92.0	2784	9	AF336981	AF336981 Homo sapi
18	18.4	92.0	2850	9	AK093058	AK093058 Homo sapi
19	18.4	92.0	2892	10	AF157105	AF157105 Mus muscu
20	18.4	92.0	2893	6	AX356951	AX356951 Sequence
21	18.4	92.0	2893	6	AX463057	AX463057 Sequence
22	18.4	92.0	2925	6	AX033272	AX033272 Sequence
23	18.4	92.0	2925	10	AF176569	AF176569 Mus muscu
24	18.4	92.0	2953	6	AX473100	AX473100 Sequence
25	18.4	92.0	2975	6	AX356955	AX356955 Sequence
26	18.4	92.0	2975	6	AX463058	AX463058 Sequence
27	18.4	92.0	25807	9	AL589746	AL589746 Human DNA
28	18.4	92.0	65558	2	AC116742	AC116742 Mus muscu
29	18.4	92.0	166462	2	AC027084	AC027084 Homo sapi
30	18.4	92.0	192000	2	AP000491	AP000491 Mus muscu
31	18.4	92.0	208249	2	AL607032	AL607032 Mus muscu
32	17.4	87.0	117019	2	AC115638	AC115638 Rattus no
33	17.4	87.0	119481	9	AC019048	AC019048 Homo sapi
34	17.4	87.0	151276	2	AC120577	AC120577 Rattus no
35	17.4	87.0	163380	2	AC113379	AC113379 Homo sapi
36	16.8	84.0	11777	1	AE001486	AE001486 Helicobac
37	16.8	84.0	112357	2	AC112746	AC112746 Rattus no
38	16.8	84.0	134028	9	AL662906	AL662906 Human DNA
39	16.8	84.0	138107	9	AC004967	AC004967 Homo sapi
40	16.8	84.0	142517	9	AC092795	AC092795 Homo sapi
41	16.8	84.0	145059	2	AC114099	AC114099 Rattus no
42	16.8	84.0	174316	2	AC117030	AC117030 Rattus no
43	16.8	84.0	174582	9	HUAC003010	AC003010 Homo sapi
44	16.8	84.0	176973	2	AC130449	AC130449 Homo sapi
45	16.8	84.0	182123	2	AC002399	AC002399 Homo sapi

ALIGNMENTS

RESULT 1  
AX014716  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AX014716  
Sequence 16 from Patent WO9953077.  
AX014716  
AX014716.1 GI:10040989  
synthetic construct.  
synthetic construct  
artificial sequences.  
1 (bases 1 to 20)  
Schwartz,J.C., Gros,C., Oulmet,T., Rose,C., Bonhomme,M.C. and  
Fachinetti,P.  
Novel nep ii membrane metalloprotease and its use for screening  
inhibitors useful in therapy

20 bp  
DNA  
linear  
PAT 07-SEP-2000

## JOURNAL

Patent: WO 9953077-A 16 21-OCT-1999;  
 INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS  
 CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME  
 MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

## FEATURES

source  
 1. 20  
 /organism="synthetic construct"  
 /db\_xref="taxon:32630"  
 /note="oligonucleotide" 2 t

## BASE COUNT

6 a 7 c 5 g

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 20;

Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

1 ATGACCAGAACTCCAGCGG 20

## Db

1 ATGACCAGAACTCCAGCGG 20

## RESULT 2

AX014701 2765 bp DNA linear PAT 07-SEP-2000

## LOCUS

Sequence 1 from Patent WO9953077

## DEFINITION

AX014701

## ACCESSION

AX014701.1 GI:10040975

## KEYWORDS

black rat.

## SOURCE

Rattus rattus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;

## REFERENCE

1 (bases 1 to 2765)

Schwartz,J.C., Gros,C., Oumet,T., Rose,C., Bonhomme,M.C. and

Facchinetti,P.

Novel nep ii membrane metalloprotease and its use for screening

inhibitors useful in therapy

Patent: WO 9953077-A 1 21-OCT-1999;

INST NAT SANTE RECH MED (FR); SCHWARTZ JEAN CHARLES (FR); GROS

CLAUDE (FR); OUMET TANJA (FR); ROSE CHRISTIANE (FR); BONHOMME

MARIE CHANTAL (FR); FACCHINETTI PATRICIA (FR)

Location/Qualifiers

1. 2765

/organism="Rattus rattus"

/db\_xref="taxon:10117"

107. 2431

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC07576.1"

/db\_xref="GI:10040976"

/translation="MGKSSVGMERADNCRRLRGVEGGLVLLVLLTLLMGAIVTL

GVPSYSGKQLPLNLSLHVSRHETVVRVLRDSSKSDICTTSCVIAARILQND

QSKPCDNFYACGGWRHIVPETNSRVSDFDLRDEVLKGVLEDSVQHRPA

LVKAKTLRSCMNQSVIEKRDSEPLNVLDMGVPAMDKNMTGPKWELEKQLAV

LNQFNRRLVLDLFINWDQNSRRHYIYIDPTLGMPSREYFKEDSHRVEAYLQFM

TSVATLRDLNLPGETDLVOEAMQVHLHLELANATVPOEKRDHYTALYHRMGLEE

LOREFGLKGNWTLFIONLVSSQVVELLPNEEVVYGYIPYLENLEEIDVFPQOTLN

YLVRVLDRIGLSQRFKEARDYRKALYGTMEEVWRRECYSVNSNREAGVSLY

IKRAFSDKSIYSELLEKRSVFDNLDELNMDESKKKAQKALNIREQIGYDY

IKEDNRHLDDEEYSSLTFSDFLFGNLGNKNAQSLKLRKREKPDNLWIGAAVY

NAPYSNRLIVFAGILQPPFFSKDQPALNFGGIGVIGHEITHGFDNNGRNFRA

GNMLDWSNRVSRHFROOSCMYQYNSFNSWELADNQCINVCFSYLGNIADNGVROA

YKAYLWLAGGRDQRLPGLNLTVAQLFFINVAQVWCGSYRPEAFIQSIKTDVHSPLK

YRVLSQNLNLPGFSEAFHCPNGSPMNRRIW"

BASE COUNT 684 a 735 c 787 g 559 t

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 2765;

Best Local Similarity 100.0%; Pred. No. 3.7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

1 ATGACCAGAACTCCAGCGG 20

## Db

813 ATGACCAGAACTCCAGCGG 832

## RESULT 3

AC094732/c

LOCUS

DEFINITION

56 unordered pieces.

ACCESSION

AC094732

VERSION

HTG: HTGS\_PHASE1

KEYWORDS

Rattus norvegicus

SOURCE

ORGANISM

Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 174953)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alzbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,

Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flaeg,N., Ford,J.,

Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,

Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,

Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,

Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,

Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,

Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,

Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Loulseged,H., Lozardo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,

Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,

Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,

Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,

Sisson,I., Sodergren,E., Sonaikhe,T., Sparks,A., Stanley,H.,

Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Tamerisa,K.,

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,

Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,

Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,

Watling,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,

Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,

Weinstock,G. and Gibbs.R.

Direct Submission

Unpublished

2 (bases 1 to 174953)

Worley,K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15624568.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBGF

Center clone name: CH230-516  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329First call to findPhrapList  
Consensus quality: 152255 bases at least Q40  
Consensus quality: 158448 bases at least Q30  
Consensus quality: 164461 bases at least Q20  
Estimated insert size: 155965; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 56 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved  
\* 1 13782: contig of 13782 bp in length  
\* 13783 13882: gap of unknown length  
\* 13883 23287: contig of 9405 bp in length  
\* 23288 23887: gap of unknown length  
\* 23888 28081: contig of 4694 bp in length  
\* 28082 28181: gap of unknown length  
\* 28182 33807: contig of 5626 bp in length  
\* 33808 33907: gap of unknown length  
\* 33908 33271: contig of 5364 bp in length  
\* 33272 39371: gap of unknown length  
\* 39372 44270: contig of 4899 bp in length  
\* 44271 44370: gap of unknown length  
\* 44371 47723: contig of 3353 bp in length  
\* 47724 47823: gap of unknown length  
\* 47824 53427: contig of 5604 bp in length  
\* 53428 53527: gap of unknown length  
\* 53528 57303: contig of 3776 bp in length  
\* 57304 57403: gap of unknown length  
\* 57404 62018: contig of 4615 bp in length  
\* 62019 62118: gap of unknown length  
\* 62119 66676: contig of 4558 bp in length  
\* 66677 66776: gap of unknown length  
\* 66777 70201: contig of 3425 bp in length  
\* 70202 70301: gap of unknown length  
\* 70302 74082: contig of 3781 bp in length  
\* 74083 74182: gap of unknown length  
\* 74183 78306: contig of 4124 bp in length  
\* 78307 78406: gap of unknown length  
\* 78407 81222: contig of 2816 bp in length  
\* 81223 81322: gap of unknown length  
\* 81323 84552: contig of 3230 bp in length  
\* 84553 84652: gap of unknown length  
\* 84653 88839: contig of 4187 bp in length  
\* 88840 88939: gap of unknown length  
\* 88940 92256: contig of 3317 bp in length  
\* 92257 92356: gap of unknown length  
\* 92357 94995: contig of 2639 bp in length  
\* 94996 95095: gap of unknown length  
\* 95096 98465: contig of 3370 bp in length  
\* 98466 98565: gap of unknown length  
\* 98566 101632: contig of 3067 bp in length  
\* 101633 101732: gap of unknown length  
\* 101733 104939: contig of 3207 bp in length  
\* 104940 105039: gap of unknown length  
\* 105040 109248: contig of 4209 bp in length  
\* 109249 109348: gap of unknown length  
\* 109349 12433: contig of 3087 bp in length  
\* 112436 112535: gap of unknown length  
\* 114676 114776: contig of 2141 bp in length  
\* 114777 117567: contig of 2791 bp in length  
\* 117568 117667: gap of unknown length

\* 117668 119209: contig of 1542 bp in length  
\* 119210 119309: gap of unknown length  
\* 119310 121797: contig of 2488 bp in length  
\* 121798 121897: gap of unknown length  
\* 121898 124293: contig of 2396 bp in length  
\* 124294 124393: gap of unknown length  
\* 124394 126428: contig of 2035 bp in length  
\* 126429 126528: gap of unknown length  
\* 126529 129525: contig of 2997 bp in length  
\* 129526 129625: gap of unknown length  
\* 129626 132334: contig of 2709 bp in length  
\* 132335 132434: gap of unknown length  
\* 132435 135274: contig of 2840 bp in length  
\* 135275 135374: gap of unknown length  
\* 135375 138074: contig of 2700 bp in length  
\* 138075 138174: gap of unknown length  
\* 138175 139985: contig of 1811 bp in length  
\* 139986 140085: gap of unknown length  
\* 140086 142273: contig of 2188 bp in length  
\* 142274 142373: gap of unknown length  
\* 142374 143598: contig of 1225 bp in length  
\* 143599 143698: gap of unknown length  
\* 143699 145434: contig of 1736 bp in length  
\* 145435 145534: gap of unknown length  
\* 145535 146985: contig of 1451 bp in length  
\* 146986 147085: gap of unknown length  
\* 147086 148099: contig of 1014 bp in length  
\* 148100 148199: gap of unknown length  
\* 148200 150915: contig of 2716 bp in length  
\* 150916 151015: gap of unknown length  
\* 151016 152501: contig of 1486 bp in length  
\* 152502 152601: gap of unknown length  
\* 152602 154010: contig of 1409 bp in length  
\* 154011 154110: gap of unknown length  
\* 154111 155758: contig of 1648 bp in length  
\* 155759 155858: gap of unknown length  
\* 155859 157622: contig of 1764 bp in length  
\* 157623 157722: gap of unknown length  
\* 157723 159428: contig of 1706 bp in length  
\* 159429 159528: gap of unknown length  
\* 159529 161209: contig of 1681 bp in length  
\* 161210 161309: gap of unknown length  
\* 161310 163413: contig of 2104 bp in length  
\* 163414 163513: gap of unknown length  
\* 163514 164702: contig of 1189 bp in length  
\* 164703 164802: gap of unknown length  
\* 164803 165998: contig of 1196 bp in length  
\* 165999 166098: gap of unknown length  
\* 166099 167412: contig of 1314 bp in length  
\* 167413 167512: gap of unknown length  
\* 167513 169231: contig of 1719 bp in length  
\* 169232 169331: gap of unknown length  
\* 169332 170534: contig of 1203 bp in length  
\* 170535 170634: gap of unknown length  
\* 170635 172047: contig of 1413 bp in length  
\* 172048 172147: gap of unknown length  
\* 172148 173509: contig of 1362 bp in length  
\* 173510 173609: gap of unknown length  
\* 173610 174553: contig of 1344 bp in length.

FEATURES Location/Qualifiers

Query Match 100.0%; Score 20; DB 2; Length 174953;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
|||||  
DB 60884 ATGACCAGAACTCCAGCGG 60865

RESULT 4  
AX146976  
LOCUS AX146976 2076 bp DNA linear PAT 08-JUN-2001

0  
2  
  
0  
2  
0  
2

**11111**

**11111**

RESULT 9  
AF302075  
LOCUS  
DEFINITION  
ACCESSION

```

/db_xref=GI:14275326"/translation="MGKSGPVGVSVESAGRQGRPGFLEGGLLLLLLVTAALVALG  
VLVDARGIPEAEVESECTPCGTAAARILQMPPTEPCDDFYOFACGWLRRH  
IPTNSRYISFDVRLDEWLVKALENSTAKDPAVERKARTLYRSQMNVSVTEKRGS  
OPLLDILEYVGWPVADRWNEVTGLEWELOLALMNSQFNRRVLIDLFIMNDQNS  
SRHIIYDOTLTGMPSREYVFENGSKRVKEYAIOFMVSATILLREDANLPDSCLVO  
EDWKVLETELQAATVPQEERHDVIALHYRMGLSELQSOGFLKFDPWTFLTOTVLS  
SVKIKLPEEVVAIVGYPLONLENIITYSARTIQNLVRMLVDRIIGSLSORFKDT  
RVNRYKALFTWTVVEEYWRCEQVYSNNMENAVGSLYEAFDFGDKSMVELRIDKYR  
TYFEVTLDELGMWDESSKKKAQEKAMSIREQIGHPTVIILEETNRRLDEEYSNLFNSFSD  
FPENSQLNKFGCAOSLRKLREKVDPNLWIICAAVVNAFYSPNRNQIIVFPAGILOPPP  
FPSKEQPOLNFVCAGTICWITCHEITHGFCDNGRNFDKNGNMDWNFSFQHPRSEQESC  
MIYQIGNYSHDLADEONGVNTLGNIADIADGGYRCAYRAYLKWAEGGKDQQLPGLD  
LTEHQFFINAIAGWCWGSRPEFAIOSIKTDVHSPLKIRYLVLGSLQNLAAPADFTHCAR  
GTPMHPKERCERYW"  
2270..2635  
3'UTR          601 a   759 c   813 g   462 t       1 others  
BASE COUNT  
ORIGIN
```

```

Query Match      92.0%; Score 18.4; DB 6; Length 2636;  
Best Local Similarity 95.0%; Pred. No. 31;  
Matches         19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```

QY    1 ATCACCAGAACTCCAGCCGG 20  
| | | | | | | | | | | | | |  
Db    648 ACGCCAGAACTCCAGCCGG 667
```

```

RESULT 12  
AF302076.  
LOCUS           2652 bp mRNA linear ROD 11-JUN-2001  
DEFINITION Mus musculus neprilysin-like peptidase beta mRNA, complete cds.  
ACCESSION AF302076  
VERSION AF302076.1 GI:10505361  
KEYWORDS  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2652)  
Shirohani,K., Tsubuchi,S., Iwata,N., Takaki,Y., Horigaya,W.,  
Maruyama,K., Kiryu-Seo,S., Kiyama,H., Iwata,H., Tomita,T.,  
Iwatsubo,T. and Saido,T.C.  
Neprilysin degrades both amyloid beta peptides 1-40 and 1-42 most  
rapidly and efficiently among thiorphan- and  
phosphoramidon-sensitive endopeptidases  
J. Biol. Chem. 276 (24), 21895-21901 (2001)  
21293028  
MEDLINE 11278416  
PUBMED  
REFERENCE 2 (bases 1 to 2652)  
Shirohani,K. and Saido,T.C.  
AUTHORS Direct Submission  
TITLE Submitted (02-SEP-2000) Proteolytic Neuroscience Laboratory, Brain  
Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama  
351-0198, Japan  
LOCATION/Qualifiers Location/Qualifiers  
1..2652 /organism="Mus musculus"  
/db_xref="taxon:10090"  
25..2322 /note="endopeptidase"  
/codon_start=1  
/product="neprilysin-like peptidase beta"  
/protein_id="AAC18447.1"  
/db_xref=GI:10505362"
```

```

/translatio n="MYERAGCWCKKSPGFVEGLAWL LLLLIGAIVTLGVFYSIGKOL  
PLITSLLGH SWBERTVVKRALRDSLKS DICTPSCVIA AARIENNDOSRNP CENFY  
QTACGGRAHHVIPETPNYSYDVFD ILRDELEV ILKGLEDTSOH RPVAVEKATLYRS  
CNMQSVIEKDR SPLLSVLKMGWGP VAMDKNWTKML KELOLAVLN SQSNRVRL  
IDLFINWDQNSRREYYIIDQPI LGMSPREY FQEDNNHK YRKAYLEEFYS VATLMRLK  
DNQLSKESA VRMEA EVLEETHLANATI VPQRDHHTAL HRMDLMELO QRFGUKG  
FWNTL IQNVL SVSEVLE PPFDEYVYG IYLEND EI DSYSARYT QMTQ NVLWRVL RD  
RIGLSQRS RFKARDY VKALY GTTFVEWRREC VS VNMS SAVGSL YIKRAFSDXS
```

REFERENCE	PUBMED	11278416	2 (bases 1 to 2694)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:43:28 ; Search time 125.659 Seconds  
(without alignments)  
358.431 Million cell updates/sec

Title: US-09-647-780A-16

Perfect score: 20

Sequence: 1 atgaccagaactccagcg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	AAZ28822	Rat membrane metal
2	20	100.0	20	AAZ28810	Rat membrane metal
3	18.4	92.0	2765	AAZ28810	Nucleotide sequenc
4	18.4	92.0	2076	AAZ28810	Human metalloprote
5	18.4	92.0	2232	AAZ28810	Nucleotide sequenc
6	18.4	92.0	2262	AAZ28810	Soluble secreted e
7	18.4	92.0	2286	AAZ28810	Human protease, PR
8	18.4	92.0	2318	AAZ28810	Nucleotide sequenc
9	18.4	92.0	2340	AAZ28810	Human SEP endopept

10	18.4	92.0	2636	AAZ28822	Human neprilysin-1
11	18.4	92.0	2663	AAZ28822	Human neprilysin-1
12	18.4	92.0	2676	AAZ28822	cDNA encoding neut
13	18.4	92.0	2714	AAZ28822	Human neprilysin-1
14	18.4	92.0	2893	AAZ28822	Human SEP endopept
15	18.4	92.0	2893	AAZ28822	Human soluble secr
16	18.4	92.0	2925	AAZ28822	cDNA encoding neut
17	18.4	92.0	2953	AAZ28822	cDNA encoding nove
18	18.4	92.0	2975	AAZ28822	Human SEP cDNA inc
19	16.8	84.0	339	AAZ28822	H. pylori ORF hp3e
20	16.8	84.0	342	AAZ28822	H. pylori cellular
21	16.8	84.0	517	AAZ28822	Human foetal liver
22	16.8	84.0	517	AAZ28822	Human brain expres
23	16.8	84.0	517	AAZ28822	Human bone marrow
24	16.8	84.0	517	AAZ28822	Probe #8034 used t
25	16.4	82.0	509	AAZ28822	Human foetal liver
26	16.4	82.0	509	AAZ28822	Human brain expres
27	16.4	82.0	509	AAZ28822	Human bone marrow
28	16.4	82.0	509	AAZ28822	Probe #8518 used t
29	16.4	82.0	509	AAZ28822	Human genome-deriv
30	15.8	79.0	265	AAZ28822	Human gene signatu
31	15.8	79.0	4364	AAZ28822	Propionibacterium
32	15.8	79.0	10198	AAZ28822	N. magadaii bacter
33	15.8	79.0	4403765	AAZ28822	Mycobacterium tube
34	15.8	79.0	4411529	AAZ28822	Mycobacterium tube
35	15.4	77.0	162	AAZ28822	Human vitronectin
36	15.4	77.0	162	AAZ28822	VN derived insulin
37	15.4	77.0	251	AAZ28822	Human biallelic po
38	15.4	77.0	251	AAZ28822	Human biallelic po
39	15.4	77.0	300	AAZ28822	Human bone marrow
40	15.4	77.0	300	AAZ28822	Human genome-deriv
41	15.4	77.0	397	AAZ28822	Human nervous syst
42	15.4	77.0	446	AAZ28822	Arabidopsis thalia
43	15.4	77.0	456	AAZ28822	Human bone marrow
44	15.4	77.0	456	AAZ28822	Human genome-deriv
45	15.4	77.0	624	AAZ28822	DNA encoding novel

ALIGNMENTS

RESULT 1

AAZ28822  
ID AAZ28822 standard; DNA; 20 BP.

XX AAZ28822;

XX 01-FEB-2000 (first entry)

XX Rat membrane metalloprotease NEPII gene probe #12.

XX Rat; membrane metalloprotease; neprilysin II; NEPII; inactivation; ss;  
XX neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
XX cardiovascular disease; neurodegenerative disease; growth disorder;  
XX hypochalamic-hypophyseal axis; endocrine disorder; probe; hybridisation.

XX Synthetic.

XX Rattus rattus.

XX FR2777291-A1.

XX 15-OCT-1999.

XX 08-APR-1998; 98FR-0004389.

XX 08-APR-1998; 98FR-0004389.

XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

XX Quinet T., Gros C, Haret C, Bonhomme MC, Facchinetti P;

XX Schwartz JC;

XX WPI; 1999-593429/51.

XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease  
XX  
PS Claim 3; Page 23; 29pp; French.  
XX  
XX Sequences AA228811-228827 represent probes for detecting the rat  
CC membrane metalloprotease designated neprilysine II (NEPII) gene  
CC (AA228810). NEPII is involved in (in)activation of neuronal and hormonal  
CC peptide messengers. NEPII is used to screen for specific substrates (used  
CC to detect NEPII in cells and tissues) or inhibitors, which can also be  
CC used to detect NEPII or for treatment of disorders related to peptidergic  
CC signalling in which NEPII is involved, e.g. cardiovascular or  
CC neurodegenerative diseases; growth disorders of endocrine origin;  
CC disturbances of the hypothalamic- hypophyseal axis or endocrine  
CC disorders.  
XX  
SQ Sequence 20 BP; 6 A; 7 C; 5 G; 2 T; 0 other;  
XX  
XX Query Match 100.0%; Score 20; DB 20; Length 20;  
XX Best Local Similarity 100.0%; Pred. No. 2.4; Length 20;  
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 ATGACCAGAACTCCAGCCGG 20  
DB 1 ATGACCAGAACTCCAGCCGG 20  
  
RESULT 2  
AA228810  
ID AA228810 standard; cDNA; 2765 BP.  
XX  
XX AA228810;  
XX  
XX 01-FEB-2000 (first entry)  
XX  
XX Rat membrane metalloprotease NEPII gene.  
XX  
XX Rat; membrane metalloprotease; neprilysine II; NEPII; inactivation;  
KW neuron; hormone; peptide messenger; inhibitor; detection; disorder;  
KW cardiovascular disease; neurodegenerative disease; growth disorder;  
KW hypothalamic-hypophyseal axis; endocrine disorder; ds.  
XX  
XX Rattus rattus.  
XX  
XX FR2777291-A1.  
XX  
XX 15-OCT-1999.  
XX  
XX 08-APR-1998; 98FR-0004389.  
XX  
XX 08-APR-1998; 98FR-0004389.  
XX  
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX Oulmet T, Gros C, Haret C, Bonhomme MC, Facchinetti P;  
PI Schwartz JC;  
XX  
XX WPI; 1999-593429/51.  
DR P-PSDB; AA44177.  
XX  
XX New membrane metalloprotease NEP II, involved in proteolysis of  
PT neuronal and hormonal peptides, used to screen for inhibitors,  
PT potentially useful for treating e.g. cardiovascular disease -  
XX  
XX Claim 2; Page 12-16; 29pp; French.  
XX  
XX This sequence represents the gene for the rat membrane metalloprotease  
CC designated neprilysine II (NEPII), which is involved in (in)activation  
CC of neuronal and hormonal peptide messengers. NEPII is used to screen  
CC for specific substrates (used to detect NEPII in cells and tissues) or  
CC inhibitors, which can also be used to detect NEPII or for treatment of

CC disorders related to peptidergic signalling in which NEPII is involved,  
CC e.g. cardiovascular or neurodegenerative diseases; growth disorders of  
CC endocrine origin; disturbances of the hypothalamic-hypophyseal axis or  
CC endocrine disorders.  
XX  
SQ Sequence 2765 BP; 684 A; 735 C; 787 G; 559 T; 0 other;  
XX  
XX Query Match 100.0%; Score 20; DB 20; Length 2765;  
XX Best Local Similarity 100.0%; Pred. No. 3.5;  
XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 ATGACCAGAACTCCAGCCGG 20  
DB 813 ATGACCAGAACTCCAGCCGG 832  
  
RESULT 3  
AAF89737  
ID AAF89737 standard; DNA; 2076 BP.  
XX  
XX AAF89737;  
XX  
XX 23-JUL-2001 (first entry)  
XX  
XX Nucleotide sequence of a human metalloprotease enzyme IGS5.  
XX  
KW Metalloprotease; IGS5; infection; pain; cancer; diabetes; obesity;  
KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
KW hypotension; hypertension; urinary retention; osteoporosis;  
KW angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
KW benign prostatic hypertrophy; migraine; psychotic disorder;  
KW neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
KW neurodegenerative disease; sleep disorder; epilepsy; kidney disease;  
KW cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
KW subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
KW peripheral vascular disease; Raynaud's disease; motility disorder;  
KW gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
KW inflammation; chemotherapy induced injury; tumour invasion;  
KW immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
KW severe mental retardation; dyskinesia; Huntington's disease;  
XX Gilles de la Tourette's syndrome; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 1..2076  
XX FT /\*tag- a  
XX FT /product- "metalloprotease enzyme IGS5"  
XX  
XX WO200136610-A1.  
XX  
XX 25-MAY-2001.  
XX  
XX 17-NOV-2000; 2000WO-EP11532.  
XX  
XX 19-NOV-1999; 99EP-0203862.  
XX 19-NOV-1999; 99NL-1013616.  
XX 31-MAY-2000; 2000EP-0201937.  
XX 31-MAY-2000; 2000NL-1015356.  
XX  
XX (SOLV ) SOLVAY PHARM BV.  
XX  
XX Deleersnijder W, Wiegers R, Weske M;  
XX WPI; 2001-343815/36.  
DR P-PSDB; AAB83840.  
XX  
XX New IGS5 polypeptides useful for treating infections, pain, cancer,  
PT diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
PT hypertension, urinary retention and Parkinson's disease  
XX  
XX Claim 11; Page 5-6; 115pp; English.  
XX

XX

PA (SOLV ) SOLVAY PHARM BV.  
 XX Deleersnijder W, Wiegers R, Weske M;  
 PI WPI; 2001-343815/36.  
 XX P-PSDB; AAB93842.  
 DR  
 XX New IGS5 polypeptides useful for treating infections, pain, cancer,  
 PT diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
 PT hypertension, urinary retention and Parkinson's disease  
 XX Claim 11; Page 8-9; 115pp; English.  
 PS  
 XX The present sequence encodes a human metalloprotease enzyme designated  
 CC IGS5. IGS5 polynucleotides and polypeptides are useful for treating  
 CC infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma,  
 CC Parkinson's disease, acute heart failure, hypotension, hypertension,  
 CC urinary retention, osteoporosis, angina pectoris, myocardial infarction,  
 CC stroke, ulcers, allergies, benign prostatic hypertrophy, migraine,  
 CC psychotic and neurological disorders, autism, multiple sclerosis,  
 CC Alzheimer's disease, and other neurodegenerative diseases, sleep  
 CC disorder, epilepsy, cardiovascular diseases, arteriosclerosis, cerebral  
 CC cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral  
 CC infarction, peripheral vascular disease, Raynaud's disease, kidney  
 CC diseases, gastrointestinal disorders, motility disorders and conditions  
 CC of delayed gastric emptying, post-operative or diabetic gastroparesis,  
 CC diarrhoea, inflammations, chemotherapy induced injury, tumour invasion,  
 CC immune disorders, arthritis, endotoxin shock, sepsis, complications of  
 CC diabetes mellitus, and severe mental retardation and dyskinesias, such  
 CC as Huntington's disease or Gilles de la Tourette's syndrome.  
 XX  
 SQ Sequence 2262 BP; 520 A; 628 C; 716 G; 398 T; 0 other;  
 Query Match 92.0%; Score 18.4; DB 22; Length 2262;  
 Best Local Similarity 95.0%; Pred. No. 21;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGACCAGAACTCCAGCGG 20  
 Db 641 ACGACCAGAACTCCAGCGG 660  
 RESULT 6  
 AAD28130  
 ID AAD28130 standard; DNA; 2286 BP.  
 AC AAD28130;  
 XX  
 DT 07-MAY-2002 (first entry)  
 DE Soluble secreted endopeptidase (SEP) consensus DNA.  
 XX Human; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; secreted endopeptidase; SEP; endocrine; vasotropic; analgesic;  
 KW enzyme; gynaecological; antisense-therapy; male erectile dysfunction;  
 KW MED; female sexual dysfunction; FSD; female sexual arousal disorder;  
 KW FSD; premature ejaculation; anorgasmia; vaginismus; mouse; rat; ds.  
 XX  
 OS Homo sapiens.  
 OS Mus sp.  
 OS Rattus sp.  
 XX  
 XX Key Location/Qualifiers  
 FT misc\_feature 1664..2286  
 FT /tag= a  
 FT /note= "Encodes catalytic domain"  
 XX  
 FT WO200206492-A1.  
 XX  
 XX 24-JAN-2002.  
 XX  
 XX 16-JUL-2001; 2001WO-IB01263.  
 XX  
 XX 14-JUL-2000; 2000GB-0017387.  
 PR

XX (PFIZ ) PFIZER LTD.  
 PA (PFIZ ) PFIZER INC.  
 XX Harrow ID, Stacey P, Walsh RT, Wayman CP, Wayman CP, Phillips SC;  
 PI WPI; 2002-155042/20.  
 XX  
 DR An isolated and/or purified nucleic acid encoding a human soluble  
 XX secreted endopeptidase which is useful for treating sexual dysfunction,  
 PT for e.g. male erectile dysfunction or female sexual dysfunction such as  
 PT female sexual arousal disorder  
 XX  
 PS Disclosure; Fig 6; 167pp; English.  
 XX The invention relates to an isolated and/or purified nucleic acid  
 CC encoding a human soluble secreted endopeptidase (SEP). The anti-SEP  
 CC antibody and the compound which inhibits or selectively inhibits the  
 CC human SEP protein are useful in the manufacture of a medicament for  
 CC the prophylaxis and/or treatment of sexual dysfunction, in particular  
 CC male erectile dysfunction (MED) or female sexual dysfunction (FSD)  
 CC (e.g. female sexual arousal disorder (FSAD)). The SEP polypeptides are  
 CC also useful for treating the above disorders and other disorders such  
 CC as premature ejaculation, anorgasmia, sexual pain (e.g. vaginismus) and  
 CC hypoaffective sexual desire disorder. The present sequence is SEP consensus  
 CC DNA sequence found in human, mouse and rat.  
 XX  
 SQ Sequence 2286 BP; 565 A; 601 C; 674 G; 446 T; 0 other;  
 Query Match 92.0%; Score 18.4; DB 24; Length 2286;  
 Best Local Similarity 95.0%; Pred. No. 21;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGACCAGAACTCCAGCGG 20  
 Db 689 ACGACCAGAACTCCAGCGG 708  
 RESULT 7  
 AAD30580  
 ID AAD30580 standard; CDNA; 2318 BP.  
 XX  
 AC AAD30580;  
 XX  
 DT 21-MAY-2002 (first entry)  
 DE Human protease, PRTS-13 CDNA.  
 XX  
 XX Human; protease; PRTS-13; enzyme; gastritis; cirrhosis; Crohn's disease;  
 KW gastrointestinal disorder; autoimmune; inflammatory; cell proliferative;  
 KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;  
 KW anemia; asthma; atherosclerosis; hypertension; myocardial infarction;  
 KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;  
 KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;  
 KW Pick's disease; infertility; vitiligo; drug screening; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FT CDS 338..1651  
 FT /tag= a  
 FT /product= "Human PRTS-13 protein"  
 FT sig\_peptide 338..427  
 FT /tag= b  
 FT mat\_peptide 428..1648  
 FT /tag= c  
 FT /product= "Mature PRTS-13 protein"  
 XX  
 XX WO200208396-A2.  
 XX  
 XX 31-JAN-2002.  
 XX

PF 17-JUL-2001; 2001WO-US22397.  
 XX 21-JUL-2000; 2000US-220063P.  
 PR 28-JUL-2000; 2000US-221680P.  
 PR 04-AUG-2000; 2000US-223344P.  
 PR 11-AUG-2000; 2000US-224717P.  
 PR 16-AUG-2000; 2000US-225988P.  
 PR 23-AUG-2000; 2000US-227568P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Deleage AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;  
 PI Tribouley CM, Das D, Kallick DA, Nguyen DB, Lee EA, Khan FA;  
 PI Yue H, Au-Young J, Griffin JA, Policky JL, Ramkumar J, Yang J;  
 PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;  
 PI Sanjanwala MS, Yao MG, Burford N, Walla NK, Lal P, Lee S, Todd S;  
 XX Lo TP, Tang YT, Elliott VS, Azimzai Y, Lu Y;  
 XX WPI; 2002-206082/26.  
 DR P-PSDB; AAE19176.  
 XX New human protease polypeptide, useful in diagnosis, prevention and  
 PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,  
 PT cell proliferative, developmental, epithelial and neurological  
 PT disorders  
 XX Claim 5; Page 174-175; 182pp; English.  
 XX The invention relates to an isolated human protease polypeptide (PRTS).  
 CC PRTS protein and DNA are useful for diagnosing, treating and preventing  
 CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),  
 CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,  
 CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,  
 CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,  
 CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),  
 CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders  
 CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,  
 CC Parkinson's disease), and reproductive disorders (infertility). PRTS  
 CC protein is useful in a number of drug screening techniques and to  
 CC analyse the proteome of a tissue or cell type. PRTS DNA is useful for  
 CC creating knockin humanised animals or transgenic animals to model human  
 CC diseases, in somatic or germline gene therapy and in microarrays  
 CC utilising fluids or tissues from patients to detect altered PKIN  
 CC expression. The present sequence is human PRTS-13 cDNA.  
 XX Sequence 2318 BP; 526 A; 647 C; 735 G; 410 T; 0 other;  
 SQ  
 Query Match 92.0%; Score 18.4; DB 24; Length 2318;  
 Best Local Similarity 95.0%; Pred. No. 21;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATGACCAGAACTCCAGCCGG 20  
 Db 703 ACGACCAGAACTCCAGCCGG 722  
 RESULT 8  
 ID AAF89738 standard; DNA; 2340 BP.  
 XX AAF89738;  
 XX AAF89738;  
 XX 23-JUL-2001 (first entry)  
 XX Nucleotide sequence of a human metalloprotease enzyme IGS5.  
 XX Metalloprotease; IGS5; infection; pain; cancer; diabetes; obesity;  
 KW anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;  
 KW hypotension; hypertension; urinary retention; osteoporosis;  
 KW angina pectoris; myocardial infarction; stroke; ulcer; allergy;  
 KW benign prostatic hypertrophy; migraine; psychotic disorder;  
 KW neurological disorder; autism; multiple sclerosis; Alzheimer's disease;  
 KW neurodegenerative disease; sleep disorder; epilepsy; kidney disease;

KW cardiovascular disease; arteriosclerosis; cerebrovasospasm;  
 KW subarachnoid hemorrhage; cerebral ischemia; cerebral infarction;  
 KW peripheral vascular disease; Raynaud's disease; motility disorder;  
 KW gastrointestinal disorder; gastric emptying; gastroparesis; diarrhoea;  
 KW inflammation; chemotherapy induced injury; tumour invasion;  
 KW immune disorder; arthritis; endotoxin shock; sepsis; diabetes mellitus;  
 KW severe mental retardation; dyskinesia; Huntington's disease;  
 KW Gilles de la Tourette's syndrome; ss.  
 XX Homo sapiens.  
 XX OS  
 XX Key Location/Qualifiers  
 FT CDS 1..2340  
 FT /tag= a  
 FT /product= "metalloprotease enzyme IGS5"  
 XX PN WO200136610-A1.  
 XX OS  
 XX 25-MAY-2001.  
 XX PF 17-NOV-2000; 2000WO-EP11532.  
 XX PR 19-NOV-1999; 99EP-0203862.  
 PR 19-NOV-1999; 99NL-1013616.  
 PR 31-MAY-2000; 2000EP-0201937.  
 PR 31-MAY-2000; 2000NL-1015356.  
 XX (SOLV ) SOLVAY PHARM BV.  
 XX Deleersnijder W, Wieggers R, Weske M;  
 XX WPI; 2001-343815/36.  
 XX P-PSDB; AAB83841.  
 XX New IGS5 polypeptides useful for treating infections, pain, cancer,  
 PT diabetes, obesity, anorexia, bulimia, asthma, schizophrenia,  
 PT hypertension, urinary retention and Parkinson's disease  
 XX Claim 11; Page 6-7; 115pp; English.

The present sequence encodes a human metalloprotease enzyme designated IGS5. IGS5 polynucleotides and polypeptides are useful for treating infections, pain, cancer, diabetes, obesity, anorexia, bulimia, asthma, Parkinson's disease, acute heart failure, hypotension, hypertension, urinary retention, osteoporosis, angina pectoris, myocardial infarction, stroke, ulcers, allergies, benign prostatic hypertrophy, migraine, psychotic and neurological disorders, autism, multiple sclerosis, Alzheimer's disease, and other neurodegenerative diseases, sleep disorder, epilepsy, cardiovascular diseases, arteriosclerosis, cerebrovasospasm, subarachnoid hemorrhage, cerebral ischemia, cerebral infarction, peripheral vascular disease, Raynaud's disease, kidney diseases, gastrointestinal disorders, motility disorders and conditions of delayed gastric emptying, post-operative or diabetic gastroparesis, diarrhoea, inflammations, arthritis, endotoxin shock, sepsis, complications of immune disorders, chemotherapy induced injury, tumour invasion, diabetes mellitus, and severe mental retardation and dyskinesias, such as Huntington's disease or Gilles de la Tourette's syndrome.

SQ Sequence 2340 BP; 539 A; 649 C; 739 G; 413 T; 0 other;

Query Match 92.0%; Score 18.4; DB 22; Length 2340;  
 Best Local Similarity 95.0%; Pred. No. 21;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCCGG 20  
 Db 719 ACGACCAGAACTCCAGCCGG 738

RESULT 9  
 ABN84280  
 ID ABN84280 standard; cDNA; 2580 BP.  
 XX

AC ABN84280;  
 XX 23-SEP-2002 (first entry)  
 XX Human SEP endopeptidase coding sequence.  
 XX  
 XX SEP; endopeptidase; enzyme; human; neuropeptide Y1; receptor;  
 KW male sexual dysfunction; male erectile dysfunction; obesity;  
 KW anorexia; bulimia; vasotropic; anorectic; therapy; inhibitor; gene;  
 KW ss.  
 XX Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT misc\_feature 1..65  
 FT /\*tag= a  
 FT /note= "5' partial vector sequence"  
 FT 258..2580  
 FT /\*tag= b  
 FT /product= "SEP"  
 FT /partial  
 FT /note= "the CDS does not include a stop codon"  
 FT misc\_feature 2580  
 FT /\*tag= c  
 FT /note= "the 3' end of the sequence is missing  
 FT from Figure 8 of the specification"  
 XX  
 XX WO200247670-A1.  
 XX  
 XX 20-JUN-2002.  
 XX  
 XX 10-DEC-2001; 2001WO-IB02399.  
 XX  
 XX 15-DEC-2000; 2000GB-0030647.  
 XX 06-APR-2001; 2001GB-0008730.  
 XX 23-APR-2001; 2001GB-0009910.  
 XX 04-MAY-2001; 2001GB-0011037.  
 XX 29-JUN-2001; 2001US-0895367.  
 XX 13-JUL-2001; 2001US-0905846.  
 XX 24-AUG-2001; 2001GB-0020679.  
 XX (PFIZ ) PFIZER LTD.  
 XX (PFIZ ) PFIZER INC.  
 XX  
 XX Naylor AM, Van Der Graaf PH, Wayman CP;  
 XX  
 XX WPI: 2002-547828/58.  
 XX P-PSDB: ABB79521.  
 XX  
 XX Use of an inhibitor of neuropeptide Y in the preparation of medicament  
 XX for the treatment or prevention of male erectile dysfunction  
 XX  
 XX Disclosure; Fig 8; 179pp; English.  
 XX  
 XX The present sequence is a nucleotide sequence coding for human  
 CC SEP, a soluble secreted endopeptidase, with an additional 5'  
 CC vector partial sequence. The invention relates to the use of an  
 CC inhibitor of NPY, or an inhibitor of NPY Y1 receptor, especially an  
 CC inhibitor selective for an NPY or NPY Y1 receptor associated with  
 CC male genitalia, in the preparation of a medicament for the  
 CC treatment or prevention of male sexual dysfunction, especially male  
 CC erectile dysfunction (MED). The NPY inhibitor may be used with an  
 CC auxiliary active agent such as an SEP inhibitor. The invention  
 CC provides a SEP assay that can be used to detect candidate  
 CC inhibitors of SEP. In addition to treatment of MED, NPY inhibitors  
 CC can also be used to treat abnormal drink and food intake disorders,  
 CC such as obesity, bulimia, anorexia and metabolic disorders.  
 XX  
 XX Sequence 2580 BP; 587 A; 741 C; 794 G; 458 T; 0 other;  
 XX  
 XX Query Match 92.0%; Score 18.4; DB 24; Length 2580;  
 XX Best Local Similarity 95.0%; Pred. No. 21;  
 XX Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGACCAGAACTCCAGCGG 20  
 I |||||  
 Db 968 ACGACCAGAACTCCAGCGG 987  
 RESULT 10  
 AAF59660  
 ID AAF59660 standard; cDNA; 2636 BP.  
 XX AC AAF59660;  
 XX DT 27-APR-2001 (first entry)  
 XX Human neprilysin-like membrane metalloproteinase SNEPB cDNA.  
 XX  
 XX Human; SNEPB; neprilysin-like membrane metalloproteinase;  
 KW splice variant; alternative splicing; zinc endopeptidase family;  
 KW neuropeptide; peptide hormone; processing; metabolism; vaccine;  
 KW drug screening; renal insufficiency; hepatic disease; ischaemia; pain;  
 KW hypertension; cancer; inflammation; cardiovascular disease;  
 KW neuronal disease; pancreatic disease; prostatic disease;  
 KW respiratory disease; gene therapy; nephrotropic; vasotropic; analgesic;  
 KW cerebroprotective; hypotensive; cytostatic; antiinflammatory;  
 KW cardiovascular; hepatotropic; ss.  
 XX  
 XX Homo sapiens.  
 XX OS  
 XX EP1069188-A1.  
 XX PN  
 XX 17-JAN-2001.  
 XX PD  
 XX 15-JUL-1999; 99EP-0401767.  
 XX PF  
 XX 15-JUL-1999; 99EP-0401767.  
 XX PR  
 XX (SNFI ) SANOFI-SYNTHELABO.  
 XX PA  
 XX Jagerschmidt A, Agnel M, Culouscou J;  
 XX PI  
 XX WPI: 2001-212582/22.  
 XX DR  
 XX P-PSDB: AAB60562.  
 XX  
 XX New membrane-associated metalloproteinase SNEPa, SNEPb and SNEPc  
 PT polypeptides and polynucleotides, useful for treating e.g. acute and  
 PT chronic renal insufficiency, pain, stroke, cancer, inflammation, renal  
 PT and hepatic ischemia  
 XX  
 XX Claim 5; Page 30-33; 72pp; English.  
 XX  
 XX The invention relates to the human SNEPa, SNEPb and SNEPc proteins, and  
 CC the cDNAs encoding them. SNEPa, SNEPb and SNEPc are neprilysin-like  
 CC membrane metalloproteinases and are the products of alternative splicing.  
 CC The substrate(s) for the SNEP proteins are not as yet known, although  
 CC the neprilysin family of zinc endopeptidases play key roles in the  
 CC processing and/or metabolism of neuropeptides and peptide hormones. SNEP  
 CC nucleotides may be used as hybridisation probes for cDNA and genomic  
 CC DNA; to isolate full-length cDNAs and genomic clones encoding SNEPa,  
 CC SNEPb or SNEPc; to isolate cDNA and genomic clones of SNEP homologues;  
 CC as research reagents and material for the discovery of treatments and  
 CC diagnostics for animal and human diseases; and for chromosome  
 CC identification. The SNEP proteins may be used as immunogens to  
 CC produce antibodies immunospecific for SNEPa, SNEPb or SNEPc. Such  
 CC antibodies are used to isolate or identify clones expressing the  
 CC protein, or to purify the proteins by affinity chromatography.  
 CC SNEP proteins may also be used in screening for compounds which modulate  
 CC SNEP endopeptidase activity; and to assess enzymatic cleavage of small  
 CC molecule substrates in cells, cell-free preparations, chemical libraries  
 CC and product mixtures. The SNEP proteins (as vaccine compositions),  
 CC SNEP nucleotides, and SNEP activators or inhibitors may be used  
 CC to treat acute and chronic renal insufficiency, renal and hepatic  
 CC ischaemia, pain, stroke, hypertensive disease, cancer, inflammation, as  
 CC well as cardiovascular, neuronal, pancreatic, prostatic, renal,

CC identification. The SNEP proteins may be used as immunogens to

CC The present sequence encodes a human neutral endopeptidase

xx The invention relates to the human SNEPa, SNEPb and SNEPc proteins, and  
cc the cDNAs encoding them. SNEPa, SNEPb and SNEPc are nephrilysin-like  
cc membrane metalloproteinases and are the products of alternative splicing  
cc the substrate(s) for the SNEP proteins are not as yet known, although



Wed Jul 9 09:32:16 2003

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
July 8, 2003, 01:24:03 : Search time 27.6098 Seconds  
(without alignments)  
222.151 Million cell updates/sec  
Run on:

Title: US-09-647-780A-16  
Perfect score: 20  
Sequence: 1 atgaccagaactccagccgg 20  
Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0  
Searched: 441362 seqs, 153338381 residues 882724  
Total number of hits satisfying chosen parameters:  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Query Match		Length DB		ID		Description	
1		15.8	79.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli		
2		15.8	79.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli		
3		15.8	77.0	399	3	US-09-188-930-63	Sequence 63, Appli		
4		15.4	77.0	1330	2	US-08-933-750C-80	Sequence 80, Appli		
5		15.4	77.0	1330	2	US-09-234-613-80	Sequence 1, Appli		
6		15.4	77.0	2055	4	US-08-258-026A-17	Sequence 17, Appli		
7		15.4	75.0	540	1	US-08-955-075A1-17	Sequence 1, Appli		
8		14.8	74.0	540	5	PCT-US95-723-1	Sequence 1, Appli		
9		14.8	74.0	1862	3	US-09-434-427-1	Sequence 3, Appli		
10		14.8	74.0	1862	4	US-09-173-151A-3	Sequence 32, Appli		
11		14.8	74.0	2061	4	US-09-280-116-32	Sequence 3, Appli		
12		14.8	74.0	2514	4	US-09-770-170-3	Sequence 42, Appli		
13		14.8	74.0	3271	4	US-08-356-060A-42	Sequence 42, Appli		
14		14.8	74.0	3900	2	US-08-460-900C-42	Sequence 42, Appli		
15		14.8	74.0	3900	4	US-08-674-509B-42	Sequence 42, Appli		
16		14.8	74.0	3900	4	US-08-954-698-42	Sequence 42, Appli		
17		14.8	74.0	3900	4	US-08-957-874-42	Sequence 5, Appli		
18		14.8	74.0	4434	2	US-08-540-406-5	Sequence 5, Appli		
19		14.8	74.0	4434	3	US-08-656-055-5	Sequence 5, Appli		
20		14.8	74.0	4434	4	US-08-954-668-5	Sequence 5, Appli		
21		14.8	74.0	4434	4	US-08-918-658-5	Sequence 5, Appli		
22		14.8	74.0	4434	5	PCT-US95-13233-5	Sequence 1, Appli		
23		14.8	74.0	23673	4	US-09-773-816-1	Sequence 1, Appli		
24		14.8	74.0	4411529	4	US-09-103-840A-1	Sequence 11, Appli		
25		14.8	74.0	4411529	4	US-08-494-907-11	Sequence 11, Appli		

Sequence 3, Appli  
Sequence 1, Appli  
Sequence 3, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 107, App  
Sequence 107, App  
Sequence 6, Appli  
Sequence 3, Appli

2766 3 US-08-936-135-3  
3024 1 US-08-149-100-1  
3471 4 US-09-116-473-3  
3652 3 US-08-936-135-5  
3680 2 US-08-494-907-1  
3680 2 PCT-US96-10986-1  
3680 2 US-08-494-907-2  
5076 2 PCT-US96-10986-2  
5076 2 US-08-494-907-4  
6170 2 PCT-US96-10986-4  
6170 2 PCT-US96-10986-3  
6387 2 PCT-US96-10986-4  
6387 2 PCT-US96-10986-3  
7198 4 US-08-994-035C-4  
7198 4 US-09-395-861-4  
7208 3 US-09-166-186-107  
7208 3 US-09-313-932-107  
7208 4 US-07-950-849-6  
48 1 US-07-950-849-3  
67 1 US-07-950-849-3

ALIGNMENTS

RESULT 1  
US-09-103-840A-2/c

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION: Robert D.

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R. M.

APPLICANT: FRASER, John C.

APPLICANT: VENTER, John C.

APPLICANT: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

TITLE OF INVENTION: 24366-20007.00

FILE REFERENCE: 24366-20007.00

CURRENT FILING DATE: 1998-06-24

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA Mycobacterium tuberculosis

ORGANISM: CDC 1551

FEATURE: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

OTHER INFORMATION: represent a, t, c or g

;; FILING DATE: 10-JUN-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Resnick, David R.  
;; REGISTRATION NUMBER: 34235  
;; REFERENCE/DOCKET NUMBER: 42838  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (617) 523-3400  
;; TELEFAX: (617) 523-6400  
;; TELEX: 200291 STRE UR  
;; INFORMATION FOR SEQ ID NO: 17:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 540 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
PCT-US95-07541-17

Query Match 74.0%; Score 14.8; DB 5; Length 540;

Best Local Similarity 88.9%; Pred. No. 2e+02; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 2;

QY 3 GACCAGAACTCCAGCCGG 20  
||| ||||| |||||  
Db 362 GACCAGAACTCCAGCCAG 345

## RESULT 9

US-08-999-723-1  
; Sequence 1, Application US/08999723A  
; Patent No. 6025180  
; GENERAL INFORMATION:  
; APPLICANT: Powell, David J.  
; APPLICANT: Southan, Christopher  
; APPLICANT: Chapman, Conrad G.  
; APPLICANT: Evans, Joanne R.  
; TITLE OF INVENTION: ASP1  
; FILE REFERENCE: GH70262  
; CURRENT APPLICATION NUMBER: US/08/999,723A  
; CURRENT FILING DATE: 1997-10-06  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 1862  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-08-999-723-1

Query Match 74.0%; Score 14.8; DB 3; Length 1862;  
Best Local Similarity 88.9%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GACCAGAACTCCAGCCGG 20  
||| ||||| |||||  
Db 1180 GACCAGAACTCCAGCAGG 1197

## RESULT 10

US-09-434-427-1  
; Sequence 1, Application US/09434427  
; Patent No. 6162630  
; GENERAL INFORMATION:  
; APPLICANT: POWELL, DAVID J.  
; APPLICANT: SOUTHAN, CHRISTOPHER  
; APPLICANT: CHAPMAN, CONRAD G.  
; APPLICANT: EVANS, JOANNE R.  
; TITLE OF INVENTION: ASP1  
; FILE REFERENCE: GH-70262-D1  
; CURRENT APPLICATION NUMBER: US/09/434,427  
; CURRENT FILING DATE: 1999-11-04  
; EARLIER APPLICATION NUMBER: US 08/999,723  
; EARLIER FILING DATE: 1997-10-06  
; EARLIER APPLICATION NUMBER: UK 9626022.9  
; EARLIER FILING DATE: 1996-12-14

;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 1  
;; LENGTH: 1862  
;; TYPE: DNA  
;; ORGANISM: HOMO SAPIENS  
US-09-434-427-1

Query Match 74.0%; Score 14.8; DB 4; Length 1862;  
Best Local Similarity 88.9%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GACCAGAACTCCAGCCGG 20  
||| ||||| |||||  
Db 1180 GACCAGAACTCCAGCAGG 1197

## RESULT 11

US-09-173-151A-3  
; Sequence 3, Application US/09173151A  
; Patent No. 6326472  
; GENERAL INFORMATION:  
; APPLICANT: Timans, Jacqueline C.  
; APPLICANT: Debets, Johannes Eduard Maria  
; APPLICANT: Antonius  
; APPLICANT: Sana, Theodore R.  
; APPLICANT: Bazan, J. Fernando  
; APPLICANT: Kastelein, Robert A.  
; TITLE OF INVENTION: Human Receptor Proteins; Related Reagents and Methods  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DNAX Research Institute  
; STREET: 901 California Avenue  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/173,151A  
; FILING DATE: 14-OCT-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/065,776  
; FILING DATE: 17-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/078,008  
; FILING DATE: 12-MAR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/081,883  
; FILING DATE: 15-APR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/095,987  
; FILING DATE: 10-AUG-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/078,416  
; FILING DATE: 18-MAR-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/062,066  
; FILING DATE: 15-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ching, Edwin P.  
; REGISTRATION NUMBER: 34,090  
; REFERENCE/DOCKET NUMBER: DX0767X  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)852-9196  
; TELEFAX: (650)496-1200  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 2061 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2058  
US-09-173-151A-3

Query Match 74.0%; Score 14.8; DB 4; Length 2061;  
Best Local Similarity 88.9%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACGAGAACTCCAGCC 18  
|||||  
DB 1765 ATGACGAGTACTTCAGCC 1782

## RESULT 12

US-09-280-116-32  
Sequence 32, Application US/09280116A  
Patent No. 6331427  
GENERAL INFORMATION:

APPLICANT: Robison, Keith E.  
TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs  
FILE REFERENCE: 5800-24, 035800/176965  
CURRENT APPLICATION NUMBER: US/09/280.116A  
CURRENT FILING DATE: 1999-03-26  
NUMBER OF SEQ ID NOS: 268  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 32  
LENGTH: 2514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: aspartyl proteases  
US-09-280-116-32

Query Match 74.0%; Score 14.8; DB 4; Length 2514;  
Best Local Similarity 88.9%; Pred. No. 2.3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GACGAGAACTCCAGCCG 20  
|||  
DB 1023 GACGAGAACTCCAGCAGG 1040

## RESULT 13

US-09-770-170-3  
Sequence 3, Application US/09770170  
Patent No. 6319679  
GENERAL INFORMATION:

APPLICANT: McKnight, Steven  
APPLICANT: Gardner, Kevin  
APPLICANT: Harper, Shannon  
APPLICANT: Rutter, Jared  
APPLICANT: Michnoff, Carolyn  
APPLICANT: Amezcua, Carlos  
TITLE OF INVENTION: PAS Kinase  
FILE REFERENCE: UTSD:0761  
CURRENT APPLICATION NUMBER: US/09/770.170  
CURRENT FILING DATE: 2001-01-26  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 3271  
TYPE: DNA  
ORGANISM: Drosophila melanogaster  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (121)..(3135)  
US-09-770-170-3

Query Match 74.0%; Score 14.8; DB 4; Length 3271;  
Best Local Similarity 88.9%; Pred. No. 2.3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGACCAGAACTCCAGCC 19  
|||||  
DB 1761 TGCCGAGACTCCAGCAG 1778

## RESULT 14

US-08-356-060A-42/c  
Sequence 42, Application US/08356060A  
Patent No. 5844079  
GENERAL INFORMATION:

APPLICANT: Ingham, Phillip W.  
APPLICANT: McMahon, Andrew P.  
APPLICANT: Tabin, Clifford J.  
TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
TITLE OF INVENTION: Proteins and Uses Related Thereto  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII(text)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/356,060A  
FILING DATE: 14-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/176,427  
FILING DATE: 30-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HMI-006CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3897  
US-08-356-060A-42

Query Match 74.0%; Score 14.8; DB 2; Length 3900;  
Best Local Similarity 88.9%; Pred. No. 2.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACGAGAACTCCAGCC 18  
|||||  
DB 1403 ATGACGAGACTCCGCC 1386

## RESULT 15

US-08-460-900C-42/c  
Sequence 42, Application US/08460900C  
Patent No. 6165747  
GENERAL INFORMATION:

APPLICANT: Ingham, Phillip W.  
APPLICANT: McMahon, Andrew P.  
APPLICANT: Tabin, Clifford J.  
APPLICANT: Bumcrot, David A.  
APPLICANT: Marti-Gorostiza, Elisa  
TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing  
TITLE OF INVENTION: Proteins and Uses Related Thereto  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
STREET: One Post Office Square  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,900C  
FILING DATE: 5-JUNE-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/435,093  
FILING DATE: 4-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/356,060  
FILING DATE: 14-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/176,427  
FILING DATE: 30-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Vincent, Matthew P.  
REGISTRATION NUMBER: 36,709  
REFERENCE/DOCKET NUMBER: HMV-006.05  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 832-1000  
TELEFAX: (617) 832-7000  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3900 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3897  
US-08-460-900C-42

Query Match 74.0%; Score 14.8; DB 4; Length 3900;  
Best Local Similarity 88.9%; Pred. No. 2.4e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACGAGAACTCCAGCC 18  
|||||  
DB 1403 ATGACGAGAACTCCAGCC 1386

Search completed: July 8, 2003, 09:32:16.  
Job time : 41.6598 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 19:09:54 ; Search time 115.512 Seconds  
(without alignments)  
273.390 Million cell updates/sec

Title: US-09-647-780A-16

Perfect score: 20  
Sequence: 1 atgaccagaactccagccgg 20

Scoring table: IDENTITY\_NOC

Gapop 10.0 , Gapext 1.0

Searched: 1105431 seqs, 789497651 residues

Total number of hits satisfying chosen parameters: 2210862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18.4	92.0	2893	9 US-10-017-273A-4	Sequence 4, Appli
2	18.4	92.0	2893	10 US-09-905-846-1	Sequence 1, Appli
3	18.4	92.0	2975	9 US-10-017-273A-5	Sequence 5, Appli
4	18.4	92.0	2975	10 US-09-905-846-5	Sequence 5, Appli
5	16.8	84.0	517	10 US-09-864-761-11840	Sequence 11840, A
6	16.4	82.0	509	10 US-09-864-761-12054	Sequence 12054, A
7	16	80.0	2972	9 US-10-171-311-195	Sequence 195, App
8	15.4	77.0	300	10 US-09-864-761-31210	Sequence 31210, A
9	15.4	77.0	417	9 US-09-918-995-7854	Sequence 7854, Ap
10	15.4	77.0	446	10 US-09-770-444-776	Sequence 776, App
11	15.4	77.0	456	10 US-09-864-761-14677	Sequence 14677, A
12	15.4	77.0	461	9 US-09-918-995-2681	Sequence 2681, Ap
13	15.4	77.0	465	9 US-09-918-995-32818	Sequence 32818, A
14	15.4	77.0	910	9 US-10-158-646-15	Sequence 15, Appl
15	15.4	77.0	1330	10 US-09-840-787-80	Sequence 80, Appl
16	15.4	77.0	1582	10 US-09-880-107-3671	Sequence 3671, Ap
17	15.4	77.0	5387	9 US-10-001-873-22	Sequence 22, Appl
18	15.2	76.0	373	9 US-09-954-987B-11	Sequence 11, Appl
19	15.2	76.0	401	10 US-09-983-965-4407	Sequence 4407, Ap

ALIGNMENTS

RESULT 1  
US-10-017-273A-4  
; Sequence 4, Application US/10017273A  
; Publication No. US20030119714A1  
; GENERAL INFORMATION:  
; APPLICANT: Pfizer Inc.  
; APPLICANT: Naylor, Alasdair M.  
; APPLICANT: Van Der Graaf, Pieter H  
; APPLICANT: Wayman, Christopher P.  
; TITLE OF INVENTION: Treatment of Male Sexual Dysfunction  
; FILE REFERENCE: PC22013  
; CURRENT APPLICATION NUMBER: US/10/017,273A  
; CURRENT FILING DATE: 2001-12-12  
; PRIOR APPLICATION NUMBER: US 60/265,358  
; PRIOR FILING DATE: 2001-01-31  
; PRIOR APPLICATION NUMBER: GB 0030647.2  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: GB 0108730.3  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: GB 0120679.6  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 09/905,846  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/291,722  
; PRIOR FILING DATE: 2001-05-17  
; PRIOR APPLICATION NUMBER: US 09/895,367  
; PRIOR FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 2893  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-017-273A-4

Query Match 92.0%; Score 18.4; DB 9; Length 2893;  
Best Local Similarity 95.0%; Pred. No. 7.3;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCCGG 20

Db 903 ACGACGAACTCCAGCCGG 922

Query Match	92.0%	Score 18.4;	DB 9;	Length 2975;
Best Local Similarity	95.0%	Fred. No. 7.3;		
Matches 19; Conservative	0;	Mismatches	1;	Indels 0;
				Gaps 0;

```

RESULT 5
; US-09-864-761-11840
; Sequence 11840, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED
; TITLE OF INVENTION: GENE EXPRESSION ANAL
; FILE REFERENCE: Acomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30

```

;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 11840  
;; LENGTH: 517  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC013597.3  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.82  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.74  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.63  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.55  
US-09-864-761-11840

Query Match 84.0%; Score 16.8; DB 10; Length 517;  
Best Local Similarity 90.0%; Pred. No. 51;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACGAGACTCCAGCCGG 20  
DB 494 ATGACGAGACTCCAGCCGG 513

## RESULT 6

US-09-864-761-12054/G  
; Sequence 12054, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecm1ca-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 12054  
;; LENGTH: 509  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC013451.7  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.69  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.8  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.93  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.72  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.59  
US-09-864-761-12054

Query Match 82.0%; Score 16.4; DB 10; Length 509;  
Best Local Similarity 94.4%; Pred. No. 79;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GACCAGACTCCAGCCGG 20  
DB 68 GACCAGACTCCAGCCGTG 51

## RESULT 7

US-10-171-311-195  
; Sequence 195, Application US/10171311  
; Publication No. US20030087270A1  
; GENERAL INFORMATION:  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Chen, Yan  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Monahan, John  
; APPLICANT: Kamatkar, Shubhangi  
; APPLICANT: Glatt, Karen  
; APPLICANT: Gannavarapu, Manjula  
; APPLICANT: Hoerish, Sebastian  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; TITLE OF INVENTION: OF CERVICAL CANCER  
; FILE REFERENCE: MRI-035  
; CURRENT APPLICATION NUMBER: US/10/171,311  
; CURRENT FILING DATE: 2002-06-12  
; PRIOR APPLICATION NUMBER: US 60/298,159  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,155  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/335,936  
; PRIOR FILING DATE: 2001-11-14  
; NUMBER OF SEQ ID NOS: 238  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 195  
; LENGTH: 2972  
; TYPE: DNA  
; ORGANISM: Homo sapiens

US-10-171-311-195

Query Match 80.0%; Score 16; DB 9; Length 2972;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ACCAGAACTCCAGCGG 19

|||||

Db 1148 ACCAGAACTCCAGCGG 1163

RESULT 8

US-09-864-761-31210  
; Sequence 31210, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 31210  
; LENGTH: 300  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC02094.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.53  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.75  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.46  
; OTHER INFORMATION: EST\_HUMAN HIT: A1133435.1, EVALUATE 0.00e+00

; OTHER INFORMATION: NT HIT: g111426686, EVALUATE 0.00e+00  
; OTHER INFORMATION: SWISSPROT HIT: P04004, EVALUATE 7.00e-56  
US-09-864-761-31210

Query Match 77.0%; Score 15.4; DB 10; Length 300;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACCAGAACTCCAGCGG 20

|||||

Db 172 ACCAGAACTCCAGCGG 188

RESULT 9

US-09-918-995-7854  
; Sequence 7854, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7854  
; LENGTH: 417  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(417)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-7854

Query Match 77.0%; Score 15.4; DB 9; Length 417;  
Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACCAGAACTCCAGCGG 20

|||||

Db 206 ACCAGAACTCCAGCGG 222

RESULT 10

US-09-770-444-776/c  
; Sequence 776, Application US/09770444  
; Patent No. US20020023280A1  
; GENERAL INFORMATION:  
; APPLICANT: Gorlach, Jorn  
; APPLICANT: An, Yong-Qiang  
; APPLICANT: Hamilton, Carol M.  
; APPLICANT: Price, Jennifer L.  
; APPLICANT: Raines, Tracy M.  
; APPLICANT: Yu, Yang  
; APPLICANT: Rameaka, Joshua G.  
; APPLICANT: Page, Amy  
; APPLICANT: Matthew, Abraham V.  
; APPLICANT: Ledford, Brooke L.  
; APPLICANT: Woessner, Jeffrey P.  
; APPLICANT: Haas, William David  
; APPLICANT: Garcia, Carlos A.  
; APPLICANT: Kriker, Maja  
; APPLICANT: Slader, Ted  
; APPLICANT: Davis, Keith R.  
; APPLICANT: Allen, Keith  
; APPLICANT: Hoffman, Neil  
; APPLICANT: Hurban, Patrick  
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
; TITLE OF INVENTION: thaliana

FILE REFERENCE: 2027 (PARA-016PRV)  
CURRENT APPLICATION NUMBER: US/09/770,444  
CURRENT FILING DATE: 2001-01-26  
PRIOR APPLICATION NUMBER: 60/178,502  
PRIOR FILING DATE: 2000-01-27  
NUMBER OF SEQ ID NOS: 999  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 776  
LENGTH: 446  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(446)  
OTHER INFORMATION: n = A,T,C or G  
US-09-770-444-776

Query Match 77.0%; Score 15.4; DB 10; Length 446;

Best Local Similarity 94.1%; Pred. No. 2.5e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TGACCAGAACTCCAGCC 18  
|||||||  
DB 211 TGACCAGAACTCCAGCC 195

## RESULT 11

US-09-864-761-14677

; Sequence 14677, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

; FILE REFERENCE: Aecm1ca-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 14677  
LENGTH: 456  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC002094.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.53  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.75  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.46  
US-09-864-761-14677

Query Match 77.0%; Score 15.4; DB 10; Length 456;

Best Local Similarity 94.1%; Pred. No. 2.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACCAGAACTCCAGCCGG 20  
|||||||  
DB 369 ACCAGAACTCCAGCCGG 385

## RESULT 12

US-09-918-995-2681

; Sequence 2681, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2681

; LENGTH: 461

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)...(461)

; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-2681

Query Match 77.0%; Score 15.4; DB 9; Length 461;

Best Local Similarity 94.1%; Pred. No. 2.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGC 17  
|||||||  
DB 270 ATGACCAGAACTCCAGC 286

## RESULT 13

US-09-918-995-32818

; Sequence 32818, Application US/09918995

; Publication No. US20030073623A1

; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc.

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

; FILE REFERENCE: 20411-756

; CURRENT APPLICATION NUMBER: US/09/918,995

; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US/09/235,076

; PRIOR FILING DATE: 1999-01-20

; NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 32818  
LENGTH: 465  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)...(465)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-993-32818

Query Match 77.0%; Score 15.4; DB 9; Length 465;  
Best Local Similarity 94.1%; Pred. No. 2.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACCAGAACTCCAGCGG 20  
|||||  
Db 408 ACCAGAACTCCAGCGG 424

RESULT 14  
US-10-158-646-15  
Sequence 15, Application US/10158646  
Publication No. US20030073105A1  
GENERAL INFORMATION:  
APPLICANT: Lasek, Amy K.W.  
APPLICANT: Sornasse, Thierry  
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
FILE REFERENCE: PA-0030-1 US  
CURRENT APPLICATION NUMBER: US/10/158,646  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: 60/295,239  
PRIOR FILING DATE: 2001-05-31  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: PERL Program  
SEQ ID NO 15  
LENGTH: 910  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incyte ID No. US20030073105A1 096954.1c  
FEATURE:  
NAME/KEY: unsure  
LOCATION: 46, 93  
OTHER INFORMATION: a, t, c, g, or other  
US-10-158-646-15

Query Match 77.0%; Score 15.4; DB 9; Length 910;  
Best Local Similarity 88.9%; Pred. No. 2.3e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGACCAGAACTCCAGCGG 19  
|||||  
Db 30 TGACCAGAACTCCAGCGG 47

RESULT 15  
US-09-840-787-80/c  
Sequence 80, Application US/09840787  
Patent No. US20020058264A1  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
Hillman, Jennifer L.  
Bandman, Olga  
Shah, Purvi  
Au-Young, Janice  
Yue, Henry  
Guegler, Karl J.  
Corley, Neil C.  
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES  
NUMBER OF SEQUENCES: 98  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/840,787  
FILING DATE: 23-Apr-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/518,865  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0356 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 80:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BLADNOT03  
CLONE: 1600518  
SEQUENCE DESCRIPTION: SEQ ID NO: 80 :  
US-09-840-787-80

Query Match 77.0%; Score 15.4; DB 10; Length 1330;  
Best Local Similarity 94.1%; Pred. No. 2.2e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACCAGAACTCCAGCGG 20  
|||||  
Db 1135 ACCAGAACTCCAGCGG 1119

Search completed: July 9, 2003, 02:22:30  
Job time : 119.562 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:47:28 ; Search time 1013.95 seconds  
(without alignments)  
319.453 Million cell updates/sec

Title: US-09-647-780A-16  
Perfect score: 20  
Sequence: 1 atgaccagaactccagccgg 20

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	600	13	BI989738 4044-75 M
2	17.4	87.0	453	9	AI034466
3	17.4	87.0	596	9	AJ425022
4	17	85.0	462	10	BB796338
5	17	85.0	820	12	BG751501
6	16.8	84.0	195	10	BE241645

C 7	16.8	84.0	334	13	BM308600
C 8	16.8	84.0	432	10	AW755678
C 9	16.8	84.0	562	13	BM521528
C 10	16.8	84.0	576	10	AW688064
C 11	16.8	84.0	656	10	AV319627
C 12	16.8	84.0	744	13	BI335608
C 13	16.8	84.0	947	12	BF582934
C 14	16.8	84.0	976	13	BI519715
C 15	16.8	84.0	1005	12	BG167146
C 16	16.8	84.0	4347	11	AK019848
C 17	16.4	82.0	270	10	BQ604959
C 18	16.4	82.0	353	14	BQ339913
C 19	16.4	82.0	420	9	AA695177
C 20	16.4	82.0	433	17	AZ055663
C 21	16.4	82.0	521	14	BQ250191
C 22	16.4	82.0	527	17	AZ881214
C 23	16.4	82.0	528	14	R89243
C 24	16.4	82.0	585	17	A2110368
C 25	16.4	82.0	641	10	BE384005
C 26	16.4	82.0	641	13	BI145086
C 27	16.4	82.0	668	17	AZ957296
C 28	16.4	82.0	995	17	CNS04B83
C 29	16	80.0	294	10	AW521474
C 30	16	80.0	306	13	BM318029
C 31	16	80.0	312	9	AI724897
C 32	16	80.0	370	13	BM383171
C 33	16	80.0	377	14	BQ207838
C 34	16	80.0	394	12	BG051841
C 35	16	80.0	396	9	AA157567
C 36	16	80.0	397	10	AW746546
C 37	16	80.0	418	10	AW746473
C 38	16	80.0	422	10	BB797023
C 39	16	80.0	423	9	AI236580
C 40	16	80.0	426	13	BI099399
C 41	16	80.0	433	13	BM327354
C 42	16	80.0	441	12	BG053545
C 43	16	80.0	445	12	BG556222
C 44	16	80.0	453	9	AI857438
C 45	16	80.0	466	14	N55564

ALIGNMENTS

RESULT 1  
BI989738 600 bp mRNA linear EST 20-DEC-2001  
LOCUS 4044-75 Mouse El4.5 retina lambda ZAP II Library Mus musculus CDNA,  
DEFINITION mRNA sequence.  
ACCESSION BI989738  
VERSION BI989738.1 GI:17960740  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 600)  
AUTHORS Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers  
1. 600  
/organism="Mus musculus"  
/db\_xref="taxon:10090"

FEATURES  
source



Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Nunasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipillar sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.

## FEATURES

source

Location/Qualifiers

1. .462  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="G4D0007M04"  
/clone\_lib="RIKEN full-length enriched, Leydig cells  
CRL-2065 MTC-1 cDNA"  
/cell\_type="Leydig cells"  
/cell\_line="CRL-2065 MTC-1"  
/note="pooled cell lines; (cell\_line=CRL-1751 WEHI 164), (cell\_line=CRL-2116 JC), (cell\_line=RCB-0035 WEHI-3), (cell\_line=RCB-0464 Meth-A), (cell\_line=RCB-0545 OHTA), (cell\_line=RCB-0559 K-1 . Fl), (cell\_line=RCB-1283 B16 melanoma), (cell\_type=B cells, cell\_line=CRL-1702 WEHI 231), (cell\_type=Leydig cells, cell\_line=CRL-2065 MTC-1), (cell\_type=Nullipotent stem cell, cell\_line=CRL-2070 NE), (tissue\_type=bladder, cell\_line=RCB-0544 MFR-2), (tissue\_type=bone marrow, cell\_type=stroma cell, cell\_line=CRL-2028 SR-4987), (tissue\_type=colon, cell\_line=RCB-0549 Cle-H3), (tissue\_type=kidney, cell\_line=CRL-142 RAG), (tissue\_type=submandibular gland, cell\_line=CRL-1734 SCA-9 clone 15), (strain=BA16/C, cell\_type=B cells, cell\_line=CRL-1669 BCL1 Clone 13.20-3B3), (strain=C3H, tissue\_type=brain, cell\_line=CRL-1443 BC3H1)"

BASE COUNT 77 a 136 c 130 g 119 t  
ORIGIN

Query Match 85.0%; Score 17; DB 10; Length 462;  
Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TCACCAGACTCCAGCC 18  
|||||

DB 69 TGACCAGACTCCAGCC 53  
|||||

## RESULT 5

BG751501/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

820 bp mRNA linear EST 15-MAY-2001  
602730080F1 NIH\_MGC\_43 Homo sapiens cDNA clone IMAGE:4873962 5',  
mRNA sequence.

BG751501  
BG751501.1 GI:14062154  
EST.  
Homo sapiens

Human  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 820)

NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov

Plate: L1CMI752 row: o column: 19  
High quality sequence stop: 807.

## FEATURES

source

Location/Qualifiers

1. 820

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4873962"

/clone\_lib="NIH\_MGC\_43"

/tissue\_type="normal pigmented retinal epithelium"

/lab\_host="DM108 (phage-resistant)"

/note="Organ: eye; Vector: pORF7; Site\_1: XhoI; Site\_2:

ECORI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCACGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library. I"

BASE COUNT 216 a 192 c 223 g 189 t

ORIGIN

Query Match

Best Local Similarity

Matches 17; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

4 ACCAGAACTCCAGCGG 20

|||||

DB

710 ACCAGAACTCCAGCGG 694

|||||

RESULT 6

BG241645

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

195 bp mRNA linear EST 03-OCT-2001  
TC0A01E0655 Pediatric acute myelogenous leukemia cell (FAB M1)  
Baylor-HGSC project-TCAA Homo sapiens cDNA clone TCAAP0655, mRNA  
sequence.

BE241645  
BE241645.1 GI:9093368  
EST.  
Homo sapiens

Human  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 195)

Wei, Y., Tsang, Y.T.M., Mei, G., Ku, J.M., Ali-Osman Jr., F.R., Muzny, D.,  
Bouck, J., Gibbs, R.A. and Margolin, J.F.

**TITLE** Pediatric Leukemia cDNA Sequencing Project  
**JOURNAL** Unpublished (2000)  
**COMMENT** Contact: Dr. Judith F. Margolin  
 Texas Children's Cancer Center and Human Genome Sequencing Center  
 at Baylor College of Medicine  
 1102 Bates, MC3-3320 Houston, TX 77030, USA  
 Tel: 832-824-4536  
 Fax: 832-825-4038  
 Email: clones@tccc.org  
 Citation: Carninci, P. and Hayashizaki, Y. High efficiency  
 full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
 Seq primer: M13 primer.

**FEATURES** Location/Qualifiers  
 source  
 1..195  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="TCAAP0655"  
 /clone\_lib="Pediatric acute myelogenous leukemia cell (FAB M1) Baylor-HGSC project-TCAA"  
 /sex="male"  
 /tissue\_type="leukopheresis"  
 /cell\_type="myeloid cell"  
 /dev\_stage="pediatric 6 years"  
 /lab\_host="DH10B"  
 /note="Vector: lambda pSB; Site\_1: BamHI; Site\_2: EcoRI;  
 First strand cDNA was primed with an anchored  
 XhoI-oligo(dT) primer [5'GGAGACTCGAGCGCCGAGGAG(T)VN  
 3'; V=A,C,G; N=A,C,G,T] and then dg tailed. Second strand  
 was primed with a BamHI-dC primer  
 [5'AGAGCTCGATCGCGCGCAATAATAAT(C) 3'].  
 Double-stranded cDNA was then digested with BamHI and XhoI  
 and directionally cloned into the BamHI and SalI sites of  
 lambda pSB vector. Library went through one round of  
 normalization. Library was constructed by Wei Yu at RIKEN  
 of Japan (Carninci P, Westover A, Nishiyama Y, Ohsumi T,  
 Itoh M, Nagaoka S, Sasakini, Okazaki Y, Muramatsu M,  
 Schneider C, Hayashizaki Y. High efficiency selection of  
 full-length cDNA by improved biotinylated cap trapper.,  
 DNA Res 4: 1, 61-6, Feb 28, 1997)"  
 BASE COUNT 39 a 74 c 53 g 29 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 195;  
 Best Local Similarity 90.0%; Pred. No. 1.7e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
 I|||||I|||||I|||||I  
 Db 86 AAGACCAGAACTCCAGCGT 105

**RESULT 7**  
 BM308600/c  
 LOCUS  
 DEFINITION sak48b10.y1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:  
 Gm-cl036-6212 5', mRNA sequence.  
 ACCESSION BM308600  
 VERSION BM308600.1 GI:18040306  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

**REFERENCE** 1 (bases 1 to 334)  
**AUTHORS** Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
 R., Waterston, R. and Wilson, R.  
 Public Soybean EST Project

**TITLE**

**JOURNAL**  
**COMMENT**

Unpublished (1999)  
 Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314-286 1800  
 Fax: 314-286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: ResGen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntsville, AL 35801 For further information  
 call: (800)-533-4363 or contact: ccu@resgen.com web site:  
 www.resgen.com

Seq primer: -40RP from Gibco  
 High quality sequence stop: 273.  
 Location/Qualifiers  
 source  
 1..334

/organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="SOYBEAN CLONE ID: Gm-cl036-6212"  
 /clone\_lib="Gm-cl036"  
 /tissue\_type="somatic embryos cultured on MSD 20"  
 /lab\_host="DH10B"  
 /note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This  
 cDNA library was constructed from mRNA isolated from  
 somatic embryos (age ranging from 2 months to 9 months)  
 cultured on MSD 20. The library was prepared using the  
 Life Technologies pSuperScript cDNA library construction  
 kit. Complementary DNA was synthesized from mRNA using a  
 poly (dT) sequence with a NotI restriction site. SalI  
 linkers adapters were ligated to the blunt-ended cDNA  
 fragments followed by NotI digestion. The cDNA fragments  
 were directionally cloned into the NotI-SalI restriction  
 site of the pSPORT1 vector. The ligated cDNA fragments  
 were transformed into E.coli Electromax DH10B host cells.  
 This library was constructed in the laboratory of Dr. Lila  
 Vodkin by Anu Khanna at the University of Illinois at  
 Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

BASE COUNT 118 a 65 c 76 g 75 t  
 ORIGIN

Query Match 84.0%; Score 16.8; DB 13; Length 334;  
 Best Local Similarity 90.0%; Pred. No. 1.8e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
 I|||||I|||||I|||||I  
 Db 217 ATGCCAGAACTCCAGCTGG 198

**RESULT 8**  
 AW755678/c

LOCUS  
 DEFINITION sl06f11.y1 Gm-cl036 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-cl036-550 5', mRNA sequence.

AW755678  
 AW755678.1 GI:7685030  
 EST.  
 KEYWORDS soybean.  
 SOURCE Glycine max

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

**REFERENCE** 1 (bases 1 to 432)  
**AUTHORS** Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers  
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk  
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann  
 R., Waterston, R. and Wilson, R.  
 Public Soybean EST Project  
 Unpublished (1999)

**TITLE** Public Soybean EST Project  
**JOURNAL**  
**COMMENT** Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact via email: ccu@resgen.com  
Insert Length: 1461 Std Error: 0.00  
High quality sequence stop: 427.

# FEATURES

Location/Qualifiers  
1..432  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="GENOME SYSTEMS CLONE ID: Gm-cl036-550"  
/clone\_lib="Gm-cl036"  
/tissue\_type="somatic embryos cultured on MSD 20"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This  
cDNA library was constructed from mRNA isolated from  
somatic embryos (age ranging from 2 months to 9 months)  
cultured on MSD 20. The library was prepared using the  
Life Technologies pSuperScript cDNA library construction  
kit. Complementary DNA was synthesized from mRNA using a  
poly (dT) sequence with a NotI restriction site. SalI  
linkers adapters were ligated to the blunt-ended cDNA  
fragments followed by NotI digestion. The cDNA fragments  
were directionally cloned into the NotI-SalI restriction  
site of the pSPORT1 vector. The ligated cDNA fragments  
were transformed into E.coli Electromax DH10B host cells.  
This library was constructed in the laboratory of Dr. Lila  
Vodkin by Anu Khanna at the University of Illinois at  
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

# BASE COUNT

141 a 71 c 119 g 99 t 2 others

Query Match 84.0%; Score 16.8; DB 10; Length 432;  
Best Local Similarity 90.0%; Pred. No. 1.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
||| ||||| ||||| |||  
DB 44 ATGCCAGAACTCCAGCTGG 25

RESULT 9  
BM521528/c  
LOCUS BM521528 562 bp mRNA linear EST 15-FEB-2002  
DEFINITION sak59a04.y1 Gm-cl036 Glycine max cDNA clone SOYBEAN CLONE ID:  
Gm-cl036-7303 5', mRNA sequence.  
ACCESSION BM521528  
VERSION BM521528.1 GI:18692680  
KEYWORDS EST.  
SOURCE soybean.  
ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE 1 (bases 1 to 562)  
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna  
,A., Bolla,B., Maria,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk  
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann  
,R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)

TITLE Contact: Shoemaker R/Public Soybean EST Project  
JOURNAL Public Soybean EST Project  
COMMENT Public Soybean EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available through: ResGen, Invitrogen Corp. 2130  
South Memorial Parkway Huntsville, AL 35801 For further information  
call: (800)-533-4363 or contact: ccu@resgen.com web site:  
www.resgen.com  
Seq primer: -40RP from Gibco  
High quality sequence stop: 429.  
Location/Qualifiers  
1..562  
/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="SOYBEAN CLONE ID: Gm-cl036-7303"  
/clone\_lib="Gm-cl036"  
/tissue\_type="somatic embryos, cultured on MSD 20"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This  
cDNA library was constructed from mRNA isolated from  
somatic embryos (age ranging from 2 months to 9 months)  
cultured on MSD 20. The library was prepared using the  
Life Technologies pSuperScript cDNA library construction  
kit. Complementary DNA was synthesized from mRNA using a  
poly (dT) sequence with a NotI restriction site. SalI  
linkers adapters were ligated to the blunt-ended cDNA  
fragments followed by NotI digestion. The cDNA fragments  
were directionally cloned into the NotI-SalI restriction  
site of the pSPORT1 vector. The ligated cDNA fragments  
were transformed into E.coli Electromax DH10B host cells.  
This library was constructed in the laboratory of Dr. Lila  
Vodkin by Anu Khanna at the University of Illinois at  
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

# FEATURES

Location/Qualifiers  
1..562

/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="SOYBEAN CLONE ID: Gm-cl036-7303"  
/clone\_lib="Gm-cl036"  
/tissue\_type="somatic embryos, cultured on MSD 20"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: NotI; Site\_2: SalI; This  
cDNA library was constructed from mRNA isolated from  
somatic embryos (age ranging from 2 months to 9 months)  
cultured on MSD 20. The library was prepared using the  
Life Technologies pSuperScript cDNA library construction  
kit. Complementary DNA was synthesized from mRNA using a  
poly (dT) sequence with a NotI restriction site. SalI  
linkers adapters were ligated to the blunt-ended cDNA  
fragments followed by NotI digestion. The cDNA fragments  
were directionally cloned into the NotI-SalI restriction  
site of the pSPORT1 vector. The ligated cDNA fragments  
were transformed into E.coli Electromax DH10B host cells.  
This library was constructed in the laboratory of Dr. Lila  
Vodkin by Anu Khanna at the University of Illinois at  
Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

# BASE COUNT

192 a 99 c 141 g 130 t

Query Match 84.0%; Score 16.8; DB 13; Length 562;  
Best Local Similarity 90.0%; Pred. No. 1.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
||| ||||| ||||| |||  
DB 317 ATGCCAGAACTCCAGCTGG 298

RESULT 10  
AW688064/c  
LOCUS AW688064 576 bp mRNA linear EST 15-JUN-2000  
DEFINITION NF004A01ST1F1000 Developing stem Medicago truncatula cDNA clone  
NF004A01ST 5', mRNA sequence.  
ACCESSION AW688064  
VERSION AW688064.1 GI:7562888  
KEYWORDS EST.  
SOURCE barrel medic.  
ORGANISM Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 576)  
AUTHORS He,X.-Z., Shadle,G., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell  
,C.J., Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Dixon  
,R.A.

TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
JOURNAL Medicago truncatula stem library  
COMMENT Unpublished (2000)  
Contact: Dixon RA  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7302  
Fax: 580 221 7380  
Email: radixon@noble.org  
Insert Length: 576 Std Error: 0.00

Plate: 004 row: A column: 01  
Seq primer: TCACACAGAAACGCTATGAC.  
Location/Qualifiers  
1. .576  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF004A01S1"  
/clone\_lib="developing stem"  
/tissue\_type="stem"  
/dev\_stage="pooled developmental"  
/note="Vector: Lambda Zap; Contains a mixture of internodal stem segments"  
150 a, 80 c 168 g 178 t

BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 576;  
Best Local Similarity 90.0%; Pred. No. 1.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCCGG 20  
||| ||||| ||||| ||||| |||||

Db 456 ATTACCAGAACTCCATCCGG 437

AV319627 656 bp mRNA linear EST 24-OCT-2001  
AV319627 RIKEN full-length enriched; 13 days embryo male testis Mus  
musculus cDNA clone 6030408W06.3', mRNA sequence.  
AV319627  
AV319627.2 GI:16394888  
house mouse.  
Mus musculus

RESULT 11  
AV319627  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 656)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Tanahashi,F., Takeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
On Nov 9, 1999 this sequence version replaced gi:6289460.  
Contact: Yoshinhide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Location/Qualifiers  
1. .856  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="6030408W06"  
/clone\_lib="RIKEN full-length enriched; 13 days embryo  
male testis"  
/sex="male"  
/tissue\_type="testis"  
/dev\_stage="13 days embryo"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 5.0 and subtraction to Rot = 100.0. Second strand  
cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGAGATTCGATTAATAATTAATCCCCCCCCCC 3']. cDNA  
was cloned into the XhoI and BamHI sites. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda  
FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."  
200 a 174 c 157 g 125 t

BASE COUNT  
ORIGIN

Query Match 84.0%; Score 16.8; DB 10; Length 656;  
Best Local Similarity 90.0%; Pred. No. 2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCCGG 20  
||| ||||| ||||| ||||| |||||

Db 404 AAGACCAGAACTCCAGCCAG 423

BI335608 744 bp mRNA linear EST 30-JUL-2001  
EST00006 Turkey Lambda Zap Library Meleagris gallopavo cDNA clone  
Nte041 5', mRNA sequence.  
BI335608  
BI335608.1 GI:15030342  
EST.  
SOURCE turkey.  
ORGANISM Meleagris gallopavo  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Meleagridae;  
Meleagris.  
1 (bases 1 to 744)  
Harry,D.E., Zaitlin,D., Marini,P.J. and Reed,K.M.  
A first-generation map of the turkey genome  
Unpublished (2001)  
Contact: Reed KM  
Department of Veterinary Pathobiology.  
University of Minnesota

295 AS/VW, 1988 Fitch Ave, St. Paul, MN 55108, USA  
 Tel: 612 624 1287  
 Fax: 612 625 0204  
 Email: reedx054etc.umn.edu  
 Seq primer: TCGAAGTTCCTCCCTCACTAAAGG  
 POLYA-Yes..

# FEATURES

source  
 Location/Qualifiers  
 1..744

/organism="Meleagris gallopavo"  
 /db\_xref="taxon:9103"  
 /clone\_lib="Nte041"  
 /clone\_lib="Turkey Lambda ZAP Library"  
 /sex="unknown"  
 /tissue\_type="whole embryo"  
 /dev\_stage="24-day"  
 /note="Vector: Lambda ZAP Library ZAP; RNA was isolated from a 24-day turkey embryo using a commercially available kit (Stratagene. cDNA library was constructed using a Lambda ZAP (SK-, Stratagene)."

BASE COUNT 191 a 174 c 186 g 193 t  
 ORIGIN  
 Query Match 84.0%; Score 16.8; DB 13; Length 744;  
 Best Local Similarity 90.0%; Pred. No. 2e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
 ||||| || ||||| |||||  
 Db 177 ATGACAAGTACTCCAGCGG 196

RESULT 13  
 BF582934/c  
 LOCUS 602101384F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4224452 5',  
 DEFINITION mRNA sequence. EST 12-DEC-2000  
 ACCESSION BF582934  
 VERSION BF582934.1 GI:11656652  
 KEYWORDS EST.  
 SOURCE house mouse.

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 1..(bases 1 to 947)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov

Plate: LLAM9814 row: h column: 21  
 High quality sequence stop: 575.  
 Location/Qualifiers

# FEATURES

source  
 1..947  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4224452"  
 /clone\_lib="NCI\_CGAP\_Co24"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 215 a 221 c 235 g 276 t  
 ORIGIN

# Query Match

Best Local Similarity 84.0%; Score 16.8; DB 12; Length 947;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
 ||||| ||||| ||||| |||||  
 Db 361 ATGACCAGAACTCCAGCGG 342

# RESULT 14

BI519715  
 LOCUS 603061950T1 NIH\_MGC\_118 Homo sapiens cDNA clone IMAGE:5211204 3',  
 DEFINITION mRNA sequence.  
 ACCESSION BI519715  
 VERSION BI519715.1 GI:15344507  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1..(bases 1 to 976)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11530 row: c column: 13  
 High quality sequence start: 22  
 High quality sequence stop: 835.

# FEATURES

source  
 1..976  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5211204"  
 /clone\_lib="NIH\_MGC\_118"  
 /tissue\_type="leukocyte"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH\_MGC Library."

BASE COUNT 127 a 335 c 292 g 222 t  
 ORIGIN

# Query Match

Best Local Similarity 84.0%; Score 16.8; DB 13; Length 976;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCGG 20  
 ||||| ||||| ||||| |||||  
 Db 125 ATGACCAGAACTCCAGCGG 144

# RESULT 15

BG167146/c  
 LOCUS 602344666F1 NIH\_MGC\_89 Homo sapiens cDNA clone IMAGE:4454728 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG167146  
 VERSION BG167146.1 GI:12673849

```

KEYWORDS  EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   1 (bases 1 to 1005)
AUTHORS    NIH-MGC http://mgs.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10247 row: c column: 17
            High quality sequence stop: 652.
FEATURES   Location/Qualifiers
            1..1005
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4454728"
                /clone_lib="NIH_MGC_89"
                /tissue_type="hypernephroma, cell line"
                /lab_host="DHI0B (phage-resistant)"
                /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.3 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT 304 a 207 c 198 g 295 t 1 others
ORIGIN

Query Match      84.0%; Score 16.8; DB 12; Length 1005;
Best Local Similarity 90.0%; Pred. No. 2.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGACCAGAACTCCAGCCGG 20
    ||| ||| ||||| |||||
Db  852 ATGTCCAAACTCCAGCCGG 833

Search completed: July 8, 2003, 09:22:00
Job time : 1018.1 secs

```